(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Publication number: **0 648 840 A2**

(12) EUROPEAN PATENT APPLICATION

(21) Application number: **94115856.0**

(22) Date of filing: **07.10.94**

(51) Int. Cl.[6]: **C12N 15/53**, C12N 9/02, C07K 14/47, C12Q 1/68, C12Q 1/26, C12N 5/10, C12N 15/85

(30) Priority: **13.10.93 US 135511**
13.10.93 US 135488
13.10.93 US 135510
28.01.94 US 187453

(43) Date of publication of application:
19.04.95 Bulletin 95/16

(84) Designated Contracting States:
AT BE CH DE DK ES FR GB GR IE IT LI LU MC NL PT SE

(71) Applicant: **Northeastern Ohio Universities College of Medicine,**
4209 State Route 44
Rootstown,
Ohio 44272-0095 (US)

(72) Inventor: **Chiang, John Young Ling**
3020 Fox Burrow Drive
Stow,
Ohio 44224 (US)

(74) Representative: **Bösl, Raphael Konrad**
Hoechst AG
Zentrale Patentabteilung
Gebäude F 821
D-65926 Frankfurt am Main (DE)

(54) Cholesterol 7alpha-hydroxylase gene regulatory elements and transcription factors.

(57) DNA regulatory elements that control cholesterol 7α-hydroxylase expression are disclosed, including bile acid responsive elements. A gene construct comprising at least one CYP7 regulatory element and a reporter gene is used to transfect HepG2 cells. Confluent transfected HepG2 cells are employed in an assay to detect a compound that modulates cholesterol 7α-hydroxylase enzyme regulation. A method for screening compounds that inhibit or stimulate expression of the enzyme is provided, as well as a method for detecting and isolating transcription factors of the cholesterol 7α-hydroxylase gene. A transcription factor of 57 KDa is identified which is useful in an assay for determining regulation of CYP7 expression.

Fig. 5

EP 0 648 840 A2

High serum cholesterol is commonly associated with an increased risk of heart attack, atherosclerosis and circulatory disorders. In addition, a variety of diseases are caused by disorder of cholesterol catabolism, such as gallstone disease, atherosclerosis, hyperlipidemia and some lipid storage diseases.

The major pathway for disposal of cholesterol in the body is by secretion of cholesterol and bile acids into the gut. Bile contains free cholesterol and bile acids. The enzyme, cholesterol 7α-hydroxylase (CYP7) commits cholesterol to bile acid synthesis and catalyzes the first and rate-limiting step of bile acid synthesis in the liver. Thus, by increasing synthesis of bile acids, this enzyme plays a key role in the liver by depleting hepatic cholesterol pools, resulting in increased LDL uptake and a lowering of serum cholesterol levels.

Bile acids are physiological agents which are important in the solubilization of lipid-soluble vitamin, sterol and xenobiotics. Bile acids are synthesized exclusively in the liver and are secreted to the intestines where they are modified to secondary bile acids. Most bile acids are reabsorbed in the ileum and recirculated to the hepatocytes via the portal vein.

The feedback of bile into the liver is known to inhibit cholesterol 7α-hydroxylase and thus inhibit the overall rate of bile acid synthesis. Cholesterol 7α-hydroxylase therefore has been a subject of intense studies to elucidate the regulatory mechanisms of bile acid synthesis in the liver.

It is known that an interruption of bile acid reabsorption, such as caused by the bile sequestrant, cholestyramine, or by a bile fistula, stimulates the rate of bile acid synthesis and cholesterol 7α-hydroxylase activity in the liver. It is believed that cholesterol 7α-hydroxylase activity in the liver is regulated primarily at the gene transcriptional level by bile acids, cholesterol, hormones, diurnal rhythm and other factors.

Generally, the regulation of eukaryotic genes is thought to occur at several locations, including the promoter sequences, located upstream of the transcription start site; enhancer or repressor sequences, located upstream of the promoter; within intron sequences, non-coding sequences located between exons or coding sequence; and in 3' sequences, located downstream from the coding region. The promoter sequence is unique to each gene and is required for the accurate and efficient initiation of gene transcription. Enhancers and/or repressors regulate promoter activity and determine the level of gene transcription during development and differentiation of a particular tissue.

The promoter of most eukaryotic genes contains a canonical TATA box which binds a TFIID TATA box binding protein. TFIID complex and associated transcription activators (TAFs) interact with the basal initiation factors and RNA polymerase II to activate promoter. The transcription complex assembly and initiation are regulated by transcription factors bound to enhancer elements located in the promoter and other regions of the gene (Pugh and Tjian, J. Biol. Chem. 267, 679-682, 1992). Tissue-specific transcription factors and nuclear steroid hormone receptors are known to play an important role in the regulation of gene expression in different tissues during development and differentiation.

However, the mechanisms underlying the regulation of cholesterol 7α-hydroxylase CYP7 gene expression at the molecular level are not understood. An understanding of regulation of CYP7 gene expression would permit development of therapeutics for treating patients with defects in bile acid synthesis and cholesterol metabolism due to altered (deficient or excessive) gene expression.

In order to study the mechanism of regulation of human cholesterol 7α-hydroxylase at the molecular level, it is therefore important to determine the correct gene sequence of its coding and promoter regions. An elucidation of its gene structure and its promoter/enhancer activity is sought in order to assay for an agent that modulates cholesterol 7α-hydroxylase enzyme regulation.

Beyond knowledge of the promoter sequence, a cell line is sought that is suitable for transfecting with a CYP7 regulatory element/reporter gene construct to determine the regulatory activity of a particular promoter region. Such a cell line then could be employed in a method for screening compounds for inhibiting or stimulating CYP7 expression by its direct or indirect interaction with the regulatory region, as reported by the reporter gene.

A method for detecting and isolating the CYP7 transcription factors also is sought. Further, upon determining a transcription factor, an assay is desired to discover other endogenous factors or exogenous agents that interact directly or indirectly with the transcription factor. Such an assay is useful to determine factors or agents that modulate the activity of the transcription factor and thereby affect expression of cholesterol 7α-hydroxylase protein.

Summary of the Invention

An embodiment of the invention provides a DNA sequence that comprises at least one regulatory element of cholesterol 7α-hydroxylase expression. In an advantageous embodiment, the DNA sequence comprises at least one regulatory element of cholesterol 7α-hydroxylase expression in either rat, human or

hamster. Another embodiment of the invention provides a rat CYP7 promoter region, deposited as clone R7αB24 on January 28, 1994, at the American Type Culture Collection, ATCC, 12301 Parkland Drive, Rockville, Maryland 20852, U.S.A., under accession number ATCC 69546.

An advantageous embodiment provides a DNA sequence comprising a regulatory element of a CYP7 gene which is selected from DNA fragments in the group consisting of human CYP7 gene fragments from about -158 to about + 32, from about -3643 to about -224, and from about -223 to about +32; and rat CYP7 gene fragments in the group consisting of from about -160 to about +32, from about -3643 to about -224, and from about -224 and +32.

Another embodiment provides a DNA sequence comprising a regulatory element of the cholesterol 7α-hydroxylase (CYP7) gene selected from DNA fragments in the group consisting of from about -191 to +64 of the rat CYP7 gene, from about -252 to +3 of the hamster CYP7 gene and from about -187 to +65 of the human CYP7 gene, or functionally active parts thereof.

Another advantageous embodiment provides DNA selected from fragments of DNA identified in Table 1, columns 1-3.

Another advantageous embodiment of the invention provides a gene construct containing at least one of the foregoing regulatory elements and a reporter gene.

Another embodiment provides a method for determining whether an agent inhibits or stimulates CYP7 gene expression. Yet other embodiments provide methods for detecting, substantially isolating and using in an assay a transcription factor of the cholesterol 7α-hydroxylase gene.

Brief Description of the Tables and Drawings

Table 1 shows the regulatory elements of rat, human and hamster CYP7 gene

Tables 2, 3 and 4 show the amino acid sequences of human, rat and hamster CYP7. Table 2 shows the human amino acid sequence (molecular-weight: 57.658; length: 504 amino acids), Table 3 shows the rat amino acid sequence (molecular-weight: 56.880; length: 503 amino acids) and Table 4 shows the hamster amino acid sequence (molecular-weight: 57.444; length: 504 amino acids).

Table 5 shows the nucleotide sequence of the region of the rat CYP7 gene taken from deposit R7αB24 and indicated by arrows in Figure 1. The transcription start site "G" is located at nucleotide position 3644. Exon I (3644-3784), Exon II (5400-5640), Exon III (6348-6934) and Exon IV (7928-7997).

Table 6 shows the approximately 5.5 kb nucleotide sequence of the λHG7α26 clone indicated by arrows in Figure 2B.

Table 7 shows the approximately 2.6 kb nucleotide sequence of the λHG7α26 clone indicated by arrows in Figure 2B.

Table 8 shows the approximately 2.3 kb nucleotide sequence of the λHG7α5 clone indicated by arrows in Figure 2C.

Table 9 shows the nucleotide sequence of the region of the hamster CYP7 gene indicated by arrows in Figure 3.

Figure 1 illustrates the rat CYP7 gene map. Boxes indicate exons. The arrows indicate the region for which a nucleic acid sequence of clone R7αB24 (shown in Figure 8) now is determined.

Figures 2A, 2B and 2C provide maps of the human CYP7 gene and clones λHG7α26 and λHG7α5. Figure 2A shows the gene map of human CYP7. Figure 2B shows the gene map of the λHGYα26 clone. Figure 2C shows the gene map of the λHG7α5 clone. Heavy boxes represent exons I, II, and III. The arrows indicate regions for which nucleic acid sequences now are determined. These sequences are shown in Tables 6, 7 and 8.

Figure 3 illustrates the hamster CYP7 gene map. The arrows indicate the region for which a sequence (shown in Table 9) now is determined.

Figure 4 shows an alignment of the proximal promoter regions of rat, human and hamster CYP7 genes. The following abbreviations are used: GRE, glucocorticoid response element; LFA1, liver factor 1; HRE, steroid/thyroid hormone response element; PPRE, peroxisome proliferator response element; TGT3, TGT3 element; and LFB1, liver factor B1. Transcription start sites "G" are indicated by a "*". Translation start codons "ATG" are underlined. The numbers indicate the nucleotide positions in each gene.

Figure 5 shows a diagram indicating the positions at which transcription factors bind to the CYP7 proximal promoter. The following abbreviations are used: HNF, hepatocyte nuclear factor; TRE, thyroid hormone response element; C/EBP, liver specific enhancer binding protein; and TFIID, TATA box binding site representing general transcription complex.

Figure 6 shows the DNase I hypersensitivity sites (I, II, III and IV) in the SacI fragment of the rat CYP7 gene. Heavy boxes are exons. A 5'-probe was used for hybridization.

Figure 7 shows the effect of bile acid conjugates on the expression of cholesterol 7α-hydroxylase mRNA levels in confluent (striped block) and subconfluent (solid block) cultures of HepG2 cells, determined by Northern blot hybridization as described in Example 3.3. The endpoint of the sequenced promoter region terminates at position -3643, while the full length of this sequence rat clone is 7997 total nucleotides long.

Figure 8 shows the effect of promoter (observed in control cells), or of added thyroxine (T_4) and dexamethasone (Dex) on the transcriptional activity of cultures of confluent (A) or subconfluent (B) HepG2 cells, transiently transfected with CYP7/LUC constructs.

Figure 9 shows the effect of bile acids on transcriptional activity of CYP7/LUC constructs transiently transfected into cultures of confluent (A) or subconfluent (B) HepG2 cells, as described in Example 3.5.

Table 2

```
Met Met Thr Thr Ser Leu Ile Trp Gly Ile Ala Ile Ala Ala Cys Cys
1               5                  10                  15

Cys Leu Trp Leu Ile Leu Gly Ile Arg Arg Arg Gln Thr Gly Glu Pro
            20                  25                  30

Pro Leu Glu Asn Gly Leu Ile Pro Tyr Leu Gly Cys Ala Leu Gln Phe
        35                  40                  45

Gly Ala Asn Pro Leu Glu Phe Leu Arg Ala Asn Gln Arg Lys His Gly
    50                  55                  60

His Val Phe Thr Cys Lys Leu Met Gly Lys Tyr Val His Phe Ile Thr
65                  70                  75                  80

Asn Pro Leu Ser Tyr His Lys Val Leu Cys His Gly Lys Tyr Phe Asp
                85                  90                  95

Trp Lys Lys Phe His Phe Ala Thr Ser Ala Lys Ala Phe Gly His Arg
            100                 105                 110

Ser Ile Asp Pro Met Asp Gly Asn Thr Thr Glu Asn Ile Asn Asp Thr
        115                 120                 125

Phe Ile Lys Thr Leu Gln Gly His Ala Leu Asn Ser Leu Thr Glu Ser
    130                 135                 140
```

```
Met Met Glu Asn Leu Gln Arg Ile Met Arg Pro Pro Val Ser Ser Asn
145                 150                 155                 160

Ser Lys Thr Ala Ala Trp Val Thr Glu Gly Met Tyr Ser Phe Cys Tyr
                165                 170                 175

Arg Val Met Phe Glu Ala Gly Tyr Leu Thr Ile Phe Gly Arg Asp Leu
            180                 185                 190

Thr Arg Arg Asp Thr Gln Lys Ala His Ile Leu Asn Asn Leu Asp Asn
        195                 200                 205

Phe Lys Gln Phe Asp Lys Val Phe Pro Ala Leu Val Ala Gly Leu Pro
    210                 215                 220

Ile His Met Phe Arg Thr Ala His Asn Ala Arg Glu Lys Leu Ala Glu
225                 230                 235                 240

Ser Leu Arg His Glu Asn Leu Gln Lys Arg Glu Ser Ile Ser Glu Leu
                245                 250                 255

Ile Ser Leu Arg Met Phe Leu Asn Asp Thr Leu Ser Thr Phe Asp Asp
            260                 265                 270

Leu Glu Lys Ala Lys Thr His Leu Val Val Leu Trp Ala Ser Gln Ala
        275                 280                 285

Asn Thr Ile Pro Ala Thr Phe Trp Ser Leu Phe Gln Met Ile Arg Asn
    290                 295                 300

Pro Glu Ala Met Lys Ala Ala Thr Glu Glu Val Lys Arg Thr Leu Glu
305                 310                 315                 320

Asn Ala Gly Gln Lys Val Ser Leu Glu Gly Asn Pro Ile Cys Leu Ser
                325                 330                 335

Gln Ala Glu Leu Asn Asp Leu Pro Val Leu Asp Ser Ile Ile Lys Glu
            340                 345                 350

Ser Leu Arg Leu Ser Ser Ala Ser Leu Asn Ile Arg Thr Ala Lys Glu
        355                 360                 365

Asp Phe Thr Leu His Leu Glu Asp Gly Ser Tyr Asn Ile Arg Lys Asp
    370                 375                 380

Asp Ile Ile Ala Leu Tyr Pro Gln Leu Met His Leu Asp Pro Glu Ile
385                 390                 395                 400

Tyr Pro Asp Pro Leu Thr Phe Lys Tyr Asp Arg Tyr Leu Asp Glu Asn
                405                 410                 415

Gly Lys Thr Lys Thr Thr Phe Tyr Cys Asn Gly Leu Lys Leu Lys Tyr
            420                 425                 430

Tyr Tyr Met Pro Phe Gly Ser Gly Ala Thr Ile Cys Pro Gly Arg Leu
        435                 440                 445

Phe Ala Ile His Glu Ile Lys Gln Phe Leu Ile Leu Met Leu Ser Tyr
    450                 455                 460

Phe Glu Leu Glu Leu Ile Glu Gly Gln Ala Lys Cys Pro Pro Leu Asp
465                 470                 475                 480
```

```
Gln Ser Arg Ala Gly Leu Gly Ile Leu Pro Pro Leu Asn Asp Ile Glu
485                 490                 495

Phe Lys Tyr Lys Phe Lys His Leu
500
```

Table 3

```
Met Met Thr Ile Ser Leu Ile Trp Gly Ile Ala Val Leu Val Ser Cys
1               5                  10                  15

Cys Ile Trp Phe Ile Val Gly Ile Arg Arg Arg Lys Ala Gly Glu Pro
            20                  25                  30

Pro Leu Glu Asn Gly Leu Ile Pro Tyr Leu Gly Cys Ala Leu Lys Phe
        35                  40                  45

Gly Ser Asn Pro Leu Glu Phe Leu Arg Ala Asn Gln Arg Lys His Gly
    50                  55                  60

His Val Phe Thr Cys Lys Leu Met Gly Lys Tyr Val His Phe Ile Thr
65                  70                  75                  80

Asn Ser Leu Ser Tyr His Lys Val Leu Cys His Gly Lys Tyr Phe Asp
                85                  90                  95

Trp Lys Lys Phe His Tyr Thr Thr Ser Ala Lys Ala Phe Gly His Arg
            100                 105                 110

Ser Ile Asp Pro Asn Asp Gly Asn Thr Thr Glu Asn Ile Asn Asn Thr
        115                 120                 125

Phe Thr Lys Thr Leu Gln Gly Asp Ala Leu Cys Ser Leu Ser Glu Ala
    130                 135                 140

Met Met Gln Asn Leu Gln Ser Val Met Arg Pro Pro Gly Leu Pro Lys
145                 150                 155                 160

Ser Lys Ser Asn Ala Trp Val Thr Glu Gly Met Tyr Ala Phe Cys Tyr
                165                 170                 175

Arg Val Met Phe Glu Ala Gly Tyr Leu Thr Leu Phe Gly Arg Asp Ile
            180                 185                 190

Ser Lys Thr Asp Thr Gln Lys Ala Leu Ile Leu Asn Asn Leu Asp Asn
        195                 200                 205

Phe Lys Gln Phe Asp Gln Val Phe Pro Ala Leu Val Ala Gly Leu Pro
    210                 215                 220

Ile His Leu Phe Lys Thr Ala His Lys Ala Arg Glu Lys Leu Ala Glu
225                 230                 235                 240

Gly Leu Lys His Lys Asn Leu Cys Val Arg Asp Gln Val Ser Glu Leu
                245                 250                 255

Ile Arg Leu Arg Met Phe Leu Asn Asp Thr Leu Ser Thr Phe Asp Asp
            260                 265                 270
```

```
Met Glu Lys Ala Lys Thr His Leu Ala Ile Leu Trp Ala Ser Gln Ala
        275                 280                 285

Asn Thr Ile Pro Ala Thr Phe Trp Ser Leu Phe Gln Met Ile Arg Ser
    290                 295                 300

Pro Glu Ala Met Lys Ala Ala Ser Glu Glu Val Ser Gly Ala Leu Gln
305                 310                 315                 320

Ser Ala Gly Gln Glu Leu Ser Ser Gly Gly Ser Ala Ile Tyr Leu Asp
                325                 330                 335

Gln Val Gln Leu Asn Asp Leu Pro Val Leu Asp Ser Ile Ile Lys Glu
            340                 345                 350

Ala Leu Arg Leu Ser Ser Ala Ser Leu Asn Ile Arg Thr Ala Lys Glu
        355                 360                 365

Asp Phe Thr Leu His Leu Glu Asp Gly Ser Tyr Asn Ile Arg Lys Asp
    370                 375                 380

Asp Met Ile Ala Leu Tyr Pro Gln Leu Met His Leu Asp Pro Glu Ile
385                 390                 395                 400

Tyr Pro Asp Pro Leu Thr Phe Lys Tyr Asp Arg Tyr Leu Asp Glu Ser
                405                 410                 415

Gly Lys Ala Lys Thr Thr Phe Tyr Ser Asn Gly Asn Lys Leu Lys Cys
            420                 425                 430

Phe Tyr Met Pro Phe Gly Ser Gly Ala Thr Ile Cys Pro Gly Arg Leu
        435                 440                 445

Phe Ala Val Gln Glu Ile Lys Gln Phe Leu Ile Leu Met Leu Ser Cys
    450                 455                 460

Phe Glu Leu Glu Phe Val Glu Ser Gln Val Lys Cys Pro Pro Leu Asp
465                 470                 475                 480

Gln Ser Arg Ala Gly Leu Gly Ile Leu Pro Pro Leu His Asp Ile Glu
                485                 490                 495

Phe Lys Tyr Lys Leu Lys His
            500
```

Table 4

```
Met Met Thr Ile Ser Leu Ile Trp Gly Ile Ala Met Val Val Cys Cys
1               5                   10                  15

Cys Ile Trp Val Ile Phe Asp Arg Arg Arg Arg Lys Ala Gly Glu Pro
            20                  25                  30

Pro Leu Glu Asn Gly Leu Ile Pro Tyr Leu Gly Cys Ala Leu Lys Phe
        35                  40                  45

Gly Ser Asn Pro Leu Glu Phe Leu Arg Ala Asn Gln Arg Lys His Gly
    50                  55                  60
```

```
His Val Phe Thr Cys Lys Leu Met Gly Lys Tyr Val His Phe Ile Thr
65                  70                  75                  80

Asn Ser Leu Ser Tyr His Lys Val Leu Cys His Gly Lys Tyr Phe Asp
                85                  90                  95

Trp Lys Lys Phe His Tyr Thr Thr Ser Ala Lys Ala Phe Gly His Arg
            100                 105                 110

Ser Ile Asp Pro Asn Asp Gly Asn Thr Thr Glu Asn Ile Asn Asn Thr
        115                 120                 125

Phe Thr Lys Thr Leu Gln Gly Asp Ala Leu His Ser Leu Ser Glu Ala
    130                 135                 140

Met Met Gln Asn Leu Gln Phe Val Leu Arg Pro Pro Asp Leu Pro Lys
145                 150                 155                 160

Ser Lys Ser Asp Ala Trp Val Thr Glu Gly Met Tyr Ala Phe Cys Tyr
                165                 170                 175

Arg Val Met Phe Glu Ala Gly Tyr Leu Thr Leu Phe Gly Arg Asp Thr
            180                 185                 190

Ser Lys Pro Asp Thr Gln Arg Val Leu Ile Leu Asn Asn Leu Asn Ser
        195                 200                 205

Phe Lys Gln Phe Asp Gln Val Phe Pro Ala Leu Val Ala Gly Leu Pro
    210                 215                 220

Ile His Leu Phe Lys Ala Ala His Lys Ala Arg Glu Gln Leu Ala Glu
225                 230                 235                 240

Gly Leu Lys His Glu Asn Leu Ser Val Arg Asp Gln Val Ser Glu Leu
                245                 250                 255

Ile Arg Leu Arg Met Phe Leu Asn Asp Thr Leu Ser Thr Phe Asp Asp
            260                 265                 270

Met Glu Lys Ala Lys Thr His Leu Ala Ile Leu Trp Ala Ser Gln Ala
        275                 280                 285

Asn Thr Ile Pro Ala Thr Phe Trp Ser Leu Phe Gln Met Ile Arg Ser
    290                 295                 300

Pro Asp Ala Leu Arg Ala Ala Ser Glu Glu Val Asn Gly Ala Leu Gln
305                 310                 315                 320

Ser Ala Gly Gln Lys Leu Ser Ser Glu Gly Asn Ala Ile Tyr Leu Asp
                325                 330                 335

Gln Ile Gln Leu Asn Asn Leu Pro Val Leu Asp Ser Ile Ile Lys Glu
            340                 345                 350

Ala Leu Arg Leu Ser Ser Ala Ser Leu Asn Ile Arg Thr Ala Lys Glu
        355                 360                 365

Asp Phe Thr Leu His Leu Glu Asp Gly Ser Tyr Asn Ile Arg Lys Asp
    370                 375                 380

Asp Ile Ile Ala Leu Tyr Pro Gln Leu Met His Leu Asp Pro Ala Ile
385                 390                 395                 400
```

```
Tyr Pro Asp Pro Leu Thr Phe Lys Tyr Asp Arg Tyr Leu Asp Glu Asn
                405                 410                 415

Lys Lys Ala Lys Thr Ser Phe Tyr Ser Asn Gly Asn Lys Leu Lys Tyr
            420                 425                 430

Phe Tyr Met Pro Phe Gly Ser Gly Ala Thr Ile Cys Pro Gly Arg Leu
        435                 440                 445

Phe Ala Val Gln Glu Ile Lys Gln Phe Leu Ile Leu Met Leu Ser Tyr
    450                 455                 460

Phe Glu Leu Glu Leu Val Glu Ser His Val Lys Cys Pro Pro Leu Asp
465                 470                 475                 480

Gln Ser Arg Ala Gly Leu Gly Ile Leu Pro Pro Leu Asn Asp Ile Glu
                485                 490                 495

Phe Lys Tyr Lys Leu Lys His Leu
            500
```

Detailed Description of the Preferred Embodiments

It was found, surprisingly, that DNA fragments comprising nucleotides downstream from about -187 of the human CYP7 gene, downstream from about -191 of the rat CYP7 gene, and downstream from about -252 of the hamster CYP7 gene are regions that exert regulatory control of transcription of the human, rat and hamster CYP7 gene, respectively.

In particular, it was found that a bile acid responsive element is located within a fragment between nucleotides -160 and +32. According to the invention, a second bile acid responsive element is located in the region between nucleotides - 3643 and -224. This was shown by transfecting hepatoma Hep2G cells with promoter/reporter constructs that contain these genetic elements within the promoter region of the construct. Thereafter the transfectants were exposed, for example, to bile acids taurodeoxycholate ("TDCA") and taurochenodeoxycholate ("TCDCA") and transcriptional activity of the reporter gene was repressed. More specifically, transcriptional activity in HepG2 cells transfected with construct pLUC-3600 was repressed by about 75%. When transfecting with pLUC-224 or pLUC-160, the transcriptional activity was repressed by about 45% or about 35% respectively, (Figure 9(A)).

Advantageously, a fragment located in the region between -160 and +32 was pinpointed to interact with at least one BARP. This fragment specifically is a direct repeat without spacing, and hence was designated as "DR_0". DR_0 in the rat is TCAAGTTCAAGT, and correspondingly in the human, is CCAAGCTCAAGT. DR_0 is a bile acid responsive element (BARE) that binds to a bile acid responsive protein (BARP) factor in the nucleus of liver cells or its nuclear extracts. Accordingly, a consensus "core" nucleotide sequence that emerges from the two species of the molecule is (T or C)CAAG(T or C).

As described in Example 2.3(b), gel shift experiments detect a BARP that binds or interacts with a bile acid responsive element 7α-TRE, for both human and rat, and human and rat DR_0 element. This BARP was characterized and possesses a molecular weight of about 57,000 Daltons, with an experimental error of about + 7000 Daltons.

Additionally, a thyroid and steroid hormone responsive element is located between -3643 and -224 of the rat CYP7 gene. This was demonstrated by increased transcriptional activity of pLUC-3600 upon stimulation with 1 μM T4 and 0.1 μM dexamethasone by 2.5-fold in confluent cultures, as demonstrated by Figure 8.

According to the present invention, the term "regulatory" means a characteristic ability of a DNA fragment to exert transcriptional control of a CYP7 gene in the presence of a factor that either down-regulates the CYP7 expression, e.g., bile salts or mevinolin, or up-regulates CYP7 expression, e.g., cholestyramine, bile fistula or cholesterol. Thus, a "regulatory element" refers to a DNA fragment disclosed in accordance with this invention that has regulatory activity with respect to CYP7.

Advantageously, an embodiment of the present invention provides a bile acid responsive element of a rat CYP7 gene which are selected from the group comprising DNA fragments from about -160 and about +32, and between about - 3643 and about -224. A further embodiment comprises a bile acid responsive element of a CYP7 human gene which is selected from the group comprising fragments from about -158 to

about +32, from about -3643 to about -224, from about -223 to about +32.

Another embodiment provides that a thyroid and steroid hormone responsive element within a fragment between about -3643 and about -224 of the rat CYP7 gene.

Another embodiment of the present invention provides a regulatory element of a CYP7 gene selected from the group comprising DNA fragments, from about -191 to about +64 of the rat CYP7 gene, from about -252 to about +3 of the hamster CYP7 gene and from about -187 to about +65 of the human CYT7 gene, and regulatory DNA fragments spanning a region within these fragments (subfragments), such as fragments shown in Figure 4.

Yet another advantageous regulatory element of the rat CYP7 gene is selected from the group of DNA fragments having regulatory activity and consisting of any of the eight fragments of DNA described in the first column of Table 1. The corresponding regulatory elements of hamster and human gene are closely homologous, as shown in Figure 4, and as listed in Table 1. Thus, an advantageous human CYP7 regulatory element is selected from the group consisting of any of the fragments of DNA described in the second column of Table 1 or human 7α-TRE, while an advantageous hamster CYP7 regulatory element is similarly selected from the group consisting of any of the eight fragments of column three of the Table 1. DNA fragments which begin at about the downstream nucleotides and end at about the upstream nucleotides as recited in Table 1 are also contemplated.

In addition to a regulatory element selected from the fragments described above (comprising from about -191 to about 64 of the rat CYP7 gene, from about -252 to about 3 of the hamster CYP7 gene and from about -187 to about 65 of the human CYT7 gene, and fragments described in Table 1), it is contemplated that other substantially homologous sequences will have CYP7 regulatory activity and thus can be used as regulatory elements in accordance with this invention. Exemplary substantially homologous sequences include: substantially homologous sequences having at least about 80%, advantageously about 90% and more advantageously about 95% nucleotide sequence homology with respect to the described fragments; sequences having at least about 82%, and advantageously at least about 90%, homology between a pair of corresponding rat and hamster DNA sequences, such homology to the sequence from about -101 to about -29 of the rat CYP7 gene and the sequence from about -161 to about -86 of the hamster CYP7 gene, for example; and sequences having homology of at least about 71%, advantageously at least about 90%, between any pair of corresponding rat and human DNA sequences, for example, about -101 to about -29 of the rat CYP7 gene and the sequence from about -104 to about -30 of the human CYP7 gene.

TABLE 1

Regulatory elements of rat, human and hamster CYP7 gene		
I. Rat	II. Human	III. Hamster
(from transcript. start site)		(from start codon)
-101 to -29	-104 to -30	-161 to -86
-81 to -37	-78 to -36	-136 to -92
-161 to -127	-159 to -124	-208 to -184
-149 to -131	-147 to -128	-206 to -188
-171 to -154	-169 to -152	-228 to -211
-101 to -82	-104 to -79	-161 to -137
-73 to -56	-71 to -54	-128 to -111
-86 to -71	-89 to -68	-146 to -126
-160 to +32	-158 to +32	-
-224 to +32	-223 to +32	-
-3643 to +32	-3643 to +32	-

Further embodiments of the present invention include a recombinant construct comprising at least one of the above-mentioned regulatory elements, advantageously a fragment disclosed in Table 1. Advantageously, for example, a regulatory element can be operably attached to a structural gene encoding CYP7, or to a reporter protein. Operably attached means that the regulatory element is positioned with respect to the structural gene such that it exerts control of the transcription of the structural gene.

A construct according to the invention can be provided in a vector capable of transforming a host cell. A host cell transformed or transfected with such a vector also comprises an embodiment of this invention, as

well as a method for expressing a selected structural gene, advantageously CYP7 or a reporter gene, using host cells of this invention. Such a method of expression comprises the steps of culturing a host cell transformed with a recombinant DNA vector comprising a gene construct comprising at least one regulatory element operably attached to the selected structural gene, wherein culturing is performed in a medium that is suitable for accommodating the desired expression, and producing the gene product.

A reporter gene allows quantitative determination of gene expression in the presence of inhibitory or stimulatory compounds. A host cell transformed with a recombinant DNA vector comprising a gene construct of at least one regulatory element operably attached to the selected structural gene provides an expression system useful in a conventional method to screen a compound for its ability to inhibit or stimulate structural gene expression. Thus, an example of a screening method provides contacting the host cell with a test compound and detecting an inhibition or stimulation of gene expression. A test compound can comprise, for example, a physiological agent derived from substances endogenous to a human or, an exogenous compound.

Regulatory elements, advantageously those fragments identified in Table 1, are used to control expression of structural genes, such as the CYP7 gene, and various reporter or indicator genes. Reporter genes include, but are not limited to, E. coli β-galactosidase, galactokinase, interleukin 2, thymidine kinase, alkaline phosphatase, luciferase and chloramphenicol acetyltransferase (CAT). Those skilled in the art readily will recognize additional reporter genes.

A representative construct of regulatory element and reporter gene ("promoter/reporter construct") is made according to Example 2.6, which employs, for example, the rat regulatory element -101 to -29. Any of the other regulatory elements according to the invention, preferably those described in Table 1, can be substituted for that rat fragment -101 to -29, by using conventional genetic engineering methods. .

According to the present invention, CYP7 constructs, such as the promoter/reporter construct, are transfected into a hepatoma cell line, advantageously, human hepatoma cell line HepG2. HepG2 liver cells express cholesterol 7α-hydroxylase normally, which makes these cells good candidates for the study of CYP7 regulation. Northern blots of normal HepG2 cells that were exposed to several bile acids, including tauro- or glyco-conjugates of cholate, deoxycholate, chenodeoxycholate or ursodeoxycholate, exhibited responsive changes in CYP7 mRNA levels as compared to non-responding control cell lines that were not exposed to those bile acids.

HepG2 cell lines are useful in screening methods provided according to the present invention. By observing expression of CYP7 in HepG2 cultures transiently transfected with CYP7 promoter/reporter gene constructs, the activity of a particular promoter region can be ascertained. Further, an agent can be added to the transfectant, and its effect on transcription can be ascertained readily.

More advantageously, a host HepG2 cell line according to the present invention that is transfected with promoter/reporter gene is both "confluent" and stable. Confluent cells are defined as cells that are at least about 4 days old, preferably 5 days, relative to the initiation of transfection. Confluent cell lines alternatively can be recognized by their uniform growth pattern, where cells tend to "adhere" to one another.

Preferably, stabilized HepG2 transfectants are employed in an assay according to the invention to provide more consistent results. A transfected cell line is stabilized using known methodology, as described by Dai et al., Biochem. 32:6928 (1993).

According to the present invention, it was discovered that the age of HepG2 transfectant cultures had a significant effect on the cells' response to steroid/thyroid hormones or bile acid conjugates. Both the endogenous cholesterol 7α-hydroxylase mRNA and transcriptional activity of the CYP7 chimeric promoter/reporter gene constructs transiently transfected into HepG2 cells responded to hydrophobic bile acids in the adult phenotype only. Younger cells were much less responsive to hormones and produced no response to bile acids, possibly due to an underdeveloped or undeveloped bile acid transport system and/or an immature steroid hormone receptor system.

Results obtained by an assay method employing confluent HepG2 cells that were transiently transfected with rat promoter/reporter constructs according to the invention identified two regions in the CYP7 gene that are responsive to bile acid repression. One bile acid responsive element (BARE) is located in the highly conserved proximal region of the promoter, from nucleotide -160 to +36, while another BARE is located in the region between -224 to -3643.

The inventive regulatory elements are also useful for detecting and isolating a transcription factor of CYP7. To detect a transcription factor, a regulatory element according to the invention, advantageously an element from Table 1, is contacted with a biological sample suspected of containing a transcription factor. Binding between the fragment and a transcription factor and the step of isolating the transcription factor are accomplished by conventional methods.

For example, to isolate a transcription factor, the following steps can be employed. First, a footprinting assay is performed to determine whether a particular gene fragment, such as a regulatory element according to the invention, binds to a nuclear transcription factor. The footprinted sequence that is revealed is used to identify DNA-protein interactions by electrophoretic mobility assay (EMSA). If a band shift is detected in EMSA, the shifted sequence is confirmed by Southwestern blot. The Southwestern blot, by SDS-polyacrylamide gel electrophoresis separates nuclear proteins. A separated protein then is incubated with a shifted DNA sequence to identify a nuclear transcription factor. The DNA sequence then is used to screen an expression cDNA library for cDNA clones encoding a transcription factor. In an alternative method, a DNA fragment of the invention can be fixed to an affinity column and used to isolate a transcription factor present in nuclear extracts (See Example 2).

An identified transcription factor can be cloned and expressed in relatively high amounts and then employed in screening compounds for the ability to influence gene expression via the specific transcription factor. For example, the effect of a bile acid or its derivatives on the function of a BARP identified according to the invention is studied by a cotransfection assay. In this assay, a CYP7 promoter/luciferase construct according to the invention, advantageously pLUC-160 and an expression plasmid containing a BARP cDNA, are cotransfected into HepG2 cell cultures. Next, an investigator determines transcriptional activity of the chimeric gene constructs (by way of the reporter gene) in the presence of test agents or endogenous factors and in control cell lines. Additionally, HepG2 cells can be transfected with a BARP, so as to express it in high amounts. Then, EMSA and footprinting assays also are performed to study the activity of a BARP.

The following examples illustrate the invention and, as such, are not to be considered as limiting the invention set forth in the claims. Either human or hamster regulatory elements can be substituted for rat regulatory elements in the following examples.

Example 1: CLONING AND NUCLEOTIDE SEQUENCING OF THE CYP7 GENES

1.(A) The Rat Gene

A rat genomic library (Clontech, RL1022j) was screened with a rat cholesterol 7α-hydroxylase cDNA previously isolated by Li et al., J. Biol. Chem. 265, 12012-12019, (1990). After screening about 1 million plaque-forming units (pfu), a positive clone, λR7α2 was plaque-purified. This clone contains a 13 kb insert that spans 8 kb of the 5'-flanking region as well as the transcription region covering exons 1 through 3 and a partial exon 4 (Figure 1). The nucleotide sequencing of an 8 kb SacI fragment is shown in Table 5 and includes the 3643 bp 5'-flanking region and coding region from exon 1 to exon 4. This fragment includes about 2 kb of the 5'-upstream region, the sequence of which was published recently by the inventor (Chiang, et al., Biochim. Biophys. Acta. 1132, 337-339, 1992). Many putative regulatory elements, including liver-enriched hepatic nuclear factors (HNFs) binding sites, steroid/thyroid hormone response elements, and ubiquitous transcription factor binding motifs (NF1, OTF-1), were identified in this gene fragment.

Table 5

```
GAGCTCTACC CTTGCTCTGC TATTGTACTT TTTAATACAC AGTTCAATCA AATGTGCCAC     60
CAGAATATGC ATGCTAACAG CTGTAGTGGT TGATTTTTCT TTCTACTCTT CTGTGTGTAA    120
GACCCCATGT TTTATCAATT ATTTTTTAAT GATTTCTTTC TTCATGCATA TGTGTGGTTG    180
TCAGTGTGAG TCTGTGTGTA CAGCAGGTGC ACAGGTATCC ACAGAGGCCA GAGGTTCCCT    240
GTAACTAGAA TTACAGGCAC TTGTGAACTT TCCTGTATGG GTGCTGGGAA GCAATCTGAG    300
GTCTTCTGCA AGGGATCTTA ACCACTGACT TTCTAGCCTG CTTTGCCCAT TTCTATTTAT    360
GATGACTGGA AACTGGGCTT AGGCCTTATA TTCTCTGAGG CCAAAATCAA GTTCTTCCAA    420
ACTGCAGGAT TTATGGTCTT CTATAGTATC CCACAGAAAT GGAAAAGAAA GTGACCCATT    480
AGAGCAGTAT TAGAGTCGAA ATAAACTCAA CTTGGTATGC CAGGACTTTG GACAATAATA    540
ACCCTGTCTT TTCAGGGCAT CTATCTGTAC TGCTGCAATA GAAACTCCAC AGGTCAGGGT    600
CACAGCTGTT GTGTTTTACA CAGTGTCCCC AGGATTAGTT CAGTGCCCAC CATGCAATAG    660
GTGTCATGGT GTGTGTGTGT GTGTGTGTGC GTGTGTCGTG CTTGTGTGCA TGTGTGTGAG    720
ACACACACAC AGAGAGATAC AAAGACAGAA ACAGAAAATT AATAAAATTT TACCAACTAA    780
AATAGGGAAT TAAAGAAAAG GAGGAGAAAA AGTTGGGCAT TCAACACCAT AAAGTCCCAG    840
TACTATGCTA AGAACACCCA GCTGTCCTCA CACCCGGGCA TGAAACTTCA TGCACTGTTC    900
ATCAGAAAAT CGTTTACACA CATCCCCTTG CAGTCTACTT GTAGTTTTAA CAACTTCAGA    960
```

GAGCACTAGC ATTTCCAGCC CCAGGTTAGA AGCTTTGGTA GATGCTGTTT GCGAGCACAG 1020

GATAGCAGCA AGAAGTGGAC TTGTTAGAAG GAAAGCCAAT GCCTATGTAA CAACGAAAAC 1080

TAAGTATGAA TCTCGAATCT CCACTCTCGT GTGTCTGTGT CTCCATATAC GTGCTTGGGT 1140

GCCTGACATG GCAAGGTGTT ACAAGTAAGG GAGGAACAAG AAAAGGACAG GGTAGTGGAC 1200

ATCAGGATGA ATGCCAGCCA GGGCGACTGG AGAGAGTCTA CGCTGCTCTG AAGGTGGGTG 1260

AAGAAGACCT CAGGAAGCTT TCTGAGGCTC CGAGAGTGCT TTTCCCTTCC CATGTTGAAA 1320

CATCCTTATT TGCAGAGAAT TCCAGGTTCA TGGGAATTTG TAAAGAGAAT ACTAAGAGGC 1380

CACCTGTGGC TTCTCCTATT TTTGTCTGCT GTCATTTATG GGACAGGGTT AGAGACCTGG 1440

CTTGCTTGGC TATGAGGCTG TTGCTTCCTC GGTTACTCTG CTGTGGTTGG ATGCATTAGG 1500

GTTAGGCCCC TCAAGAGCCA TGTGTCATTT TATAAAAGCA ATATAAATAT ACTTAAGGTG 1560

CACAAAGCAT TAGGAGGTCT GAGATAATAG ATTCTGAGAA AATCTATCCT GCTGTGTAGC 1620

AACTGATGTT TATGATTATA GTCCCAGACC ACACGATAAA GGATCTGTGG ACTCTGTTTA 1680

GGGAGGTCAA AAAACTATTG CAAATGGAGT CTATAGAGAA AACTAGACAG GACTCAATGC 1740

TCACCAATCG AGAATTAGTT GATGAGCTGG GGTAGTGACT TAGTGGATAA GAACACGGTC 1800

CTTTCAGAGG TCCTGAGTTA AATCCCCAGC AAACACATGG TGGCTCATAA CCATCTATAT 1860

TGTGATTTGA TGCCCTCTTC TGGCATGCAG GTGTACATGC AGACTCGTAT ACATAAAATA 1920

AATAAATCTT GAAAAAATGA ATACGTTGAA TAAGTGTCCC CTCGGATAAC TTTCTGCAGA 1980

ATTTTAAGCA CATGTCAATG GTAATAACAC ACACACACAC ACACACACAC ACACACACAC 2040

ACACACATAC ACACACCATA CAGATATGTA TCTAGAGACA TACACATGTA CATTTTATCT 2100

CTTTTATTTT CTTCTCCCCT CTTTGACATC AAGGAATAGA ATGCACTCAC TGTGGCCTAG 2160

TGCCACACTC TACCTATTTC TTTGGCTTTA CTTTGTGCTA GGTGACCCGA AAGGTTTAAA 2220

TATCAAAAAT GCTAATGGCT CGACATTTAC ATCCCCAATT TCTCCTTTCT CCTTACCTCA 2280

GACTCTTACA TTCAGTTGAC AATTTGACAT CGTCTCCTGG ATTTTCAAAT GTTCAGCACA 2340

CTGTACTGAT GTACTGCCTT CCAAGGCAAC CGGCACGATC CTCTCCCCAC TCCCAAGCAT 2400

CCCTCCATGA GCCAGTGTTT GCTTATCTTC TTGACTCTTG TTTTAACCCA ACTCCTCCCC 2460

TATTCACTCT GCTCTAATTC ATTCATTCTA TATTTTCGCA CATCAGGCTC ATCCTTTGCT 2520

CAGGAACTTC ACTTTTGCTT TCCGGTCTCC TGGAAATGTG TTTTCTTGGC TATTCCATCT 2580

CAAGACCATC TTTTCAGAAA AGCTTTTCCT ATCAACATAT TTAAAGCCCT CTTCATCCCC 2640

CAGTAGCTCT GGACACCTCA TTTTATGGAT ACACAACACA TATTTGCCAC CTGTCTCCCC 2700

ATTAAAATAT AATCTTCAGT AGAGAAACTC CATATCTTGT TAATACCTGA AACAAGAATA 2760

TCTTCAAAGA GTTCCTGGGA CATAAAAACG CTCAATTAAT ATTTATGTTA AACAGGGATC 2820

TGGGGTATAT CACAGAGGTA GAGGGCTTAC CTAGGAGGAG TTGGGCCATG GGTTCAACTT 2880

CCAGCACAGA ATGAAAGATT ATGTTAAATA AAGTTGGGAA GGATGTATGC CAGTCTATGA 2940

GTAGTATAGG AGGTAAATTA TGAATTCATA TTTACTTTTC GGACAAGAAG TGTTGTAGTC 3000

TTTATTTGAA ATAAAATACA TCTTAATTAC CAATAACAAT TGGTAAGGAG TGAATTCTCA 3060

AGCTGTGGCT TCCTGGTAGA TGAGTCCTGG GAGGTTTTCT ATTTCGATGA TGGTAGATAG 3120

GTAACCTGTC ATATACCACA TGAAATACCT GTGGCTTTGT AAACACACCG AGCAGTCAAG 3180

CAGGAGAATA GTTCCATACA GTTCGCGTCC CTTAGGATTG GTTTCGGGAT ACTTCTGGAG 3240

GTTCATTTAA ATAATTTTCC CCGAAGTACA TTATGGGCAG CCAGTGTTGT GATGGGAAGC 3300

TTCTGCCTGT TTTGCTTTGC GTCGTGCTCC ACACCTTTGA CAGATGTGCT CTCATCTGTT 3360

TACTTCTTTT TCTACACACA GAGCACAGCA TTAGCTGCTG TCCCGGCTTT GGATGTTATG 3420

TCAGCACATG AGGGACAGAC CTTCAGCTTA TCGAGTATTG CAGCTCTCTG TTTGTTCTGG 3480

AGCCTCTTCT GAGACTATGG ACTTAGTTCA AGGCCGGGTA ATGCTATTTT TTTCTTCTTT 3540

TTTCTAGTAG GAGGACAAAT AGTGTTTGCT TTGGTCACTC AAGTTCAAGT TATTGGATCA 3600

TGGTCCTGTG CACATATAAA GTCTAGTCAG ACCCACTGTT TCGGGACAGC CTTGCTTTGC 3660

TAGGCAAAGA GTCTCCCCTT TGGAAATTTT CCTGCTTTTG CAAAATGATG ACTATTTCTT 3720

TGATTTGGGG AATTGCCGTG TTGGTGAGCT GTTGCATATG GTTTATTGTT GGAATAAGGA 3780

GAAGGTATGG AAAGATTTTT AAAAATTTGT CTTTTAGCTT ATTTCTAGTA TTCATTGCCT 3840

TCACTATTAT GTAGTGCAAA AAATACTAAT GCATTAATAT TTTTAAATTT AAAATTTAAA 3900

GACGTACTTC TTTGACTAAA TCTAGTAAGA TGTAGAGAGT CCCCCTTGGA ACATTCACAT 3960

ATGCCACTGG TAATGCAGAT CTTGTGAAAT ATAACTAAAG AAATCACAAG TCATCGATGT 4020

AAGTTTGTGT CTGCATGGGC GGAACAAACC TAAGCTAAGA AGAGTAGTAT TTGGGAGGGA 4080

TCTTTCTGTG ACATGAACTG AATAGACGCA CTGCCTCAGC AAACACACAT TCATTTGAAT 4140

TTTCCTCAGA CTCAGTCTAA GCCTGGTGAG AGCACCAAGT GTGAGTCTGT CTGCCACTAA 4200

CGTTTCCTTC CAGTGGTAAT CAGCTGTGTG GCTGTGAAAC CTTGGCGCCT GCACATGACA 4260

GCCATTTGAA TAGTTCAAAG AACATTTAGG GACAGGATAT TAAGATATTT TCTGTGATGT 4320

CAACATCAAA ATAGGAGAAT GCCCCTGGCA TTATCTTCAG AGAGGTAGAC TACTGTGCGT 4380

TGTCTTACTT TAAAGAAATT TCTTTGCCCC TTTGGCTATT TTAATTCAAA CCTGAAAGTT 4440

TTCAGTTTTA ATTAAACTGT TGATTTTCAT GCTAGGAAAG GAAATATCAA TTATACTTAA 4500

TTGTTCTTAC AAGAAATAAA ATCATTTATG TCGGGAGATA AATAAGCTCA TAATTTTAAT 4560

AAAACATTTA AGAGAGAGAA AAAGAGTAGT GGATTATAGT TCATTGTCTG TCAATGTTTA 4620

CCTGACCCAG TTTCATTTTA TAATTATCTA ATTTTTCAAA TGAGATTCCT GTTCTTTCCA 4680

AATATCATTG CAGAATACTA ACATTCTTTT TTTCAGAGTT GAGAATCAAA TGGAGGGTTT 4740

TTTCATCCTG GCACAAGCTC CGCTCTTCAG TAACACCTCC AGCCCTCAGA ATGCCAATAT 4800

TTTAAATTAT GTAGGTTGTT AAAACTTTAG TGCTGGGGCT GGGGATTTAG CTCAGTGGTA 4860

GAGCACTTGC CTAGCAAGCG CAAGGCCCTG GGTTCGGTCC CCAGCTCTGA AAAAAAGAAA 4920

AAGAAAAAAA AAAACTTTAG TGCTGTAGCC CTTTCTGTTA TTTGATGTTT CACATCTGTT 4980

AAAAAACAAA ACAAAACAAA AAAAACAAGC AAATGGAACA TTTTAGGCAT TCTTTGGGGG 5040

AAATGATTCT TAGAGCAAGT CTAATCATTA GGTGATAGTT TCATTTTTAC ACCAAGAACA 5100

AGAATCTTGT TGGCTGTGTT AACACTTTAA GCCCTGTTGT AGGGAAAAAG CAATCAGACA 5160

CAGGCACAGA AAAGAATTTG GATGAGTACT TGATGATGTA TGTATATATG GTGAATAGAC 5220

TGATGGGTGG GCTGCTGGCT GGGTTGGTAA GTGGGTAGAT TTTTTTTTAA AGATTTATTC 5280

ATTTATTATA TATCAGTACA CTGTAGCTAT CTTCAGATAC ACCAGAAGGG CATCGGATCT 5340

CTTTACAGAT GGTTGTGAGC CACCATGTTT TCCTAACCTC TCAAGTCTCT GTCTTCCAGG 5400

AAAGCTGGTG AACCTCCTTT GGAGAACGGG TTGATTCCGT ACCTGGGCTG TGCTCTGAAA 5460

TTTGGATCTA ATCCTCTTGA GTTCCTAAGA GCTAATCAAA GGAAGCATGG TCACGTTTTT 5520

ACCTGCAAAC TGATGGGGAA ATATGTCCAT TTCATCACAA ACTCCCTGTC ATACCACAAA 5580

GTCTTATGTC ATGGAAAATA TTTTGACTGG AAAAAATTTC ATTACACTAC TTCTGCGAAG 5640

GTAATTAATT CGTTATACAG ATTCTGTTTG TTTCCTGGTC TGTTGATGTA TTAGTGTATT 5700

TAGTTGTTCC AATTTTGTTA GGTTGCAGAA TAGAGGTAAC ATAAAATCAG GGCGTTTCTT 5760

AGTAATAAGC ATTAGACATT TAAGGCAGAT GTAAACCTGT CATTGATGAT TCCGGAGACA 5820

GAGGACACTG CAGGAATCAG GAAGGTACAG ATTCATAGCA CCACTCGTCC CTTAACAACA 5880

CCCTGAGCAG GGTGTTGGCA CTCTTAGCCT TCAGTCCTTG TACACACGTT TCATTCCTAA 5940

GATATAGGCT GTATATTTAA ACACGATTTG GAAGCCATCA AGAATCTGTT CTAGAGAAAA 6000

CAGCATTTAA TGATCTTTTG CAAGAAAATA TCAGTTATAG TCTCTGTCAT TAAGTACATT 6060

GTAATCTGGT TAAAGAGTAT CTACTAAGAA AGTAAAGGCA GATTAGAACA ATACCAATGG 6120

ATGATGGGCC ATCCAGAGAA ATCCTACTGT AAATGCTGGG ATTTAAACTT GACCCCAAGG 6180

AAGAGTATGA CTTGATTCTA CCTTTGGAAT GTGCTGTAAA ATCATATTAG GGAAGGTTCC 6240

AGACAGAGAA GTGGGATGTA TTTAATCTAT CTTCCAGCCC ACTCTCTAAC ACTAGCTAGC 6300

TTTGGGCTTT AGACCCTCCC CATTTCATGG ATTCTATTTT CTACCAGGCA TTTGGACACA 6360

GAAGCATTGA CCCAAATGAT GGAAATACCA CGGAAAATAT AAACAACACT TTTACCAAAA 6420

CCCTCCAGGG AGATGCTCTG TGTTCACTTT CTGAAGCCAT GATGCAAAAC CTCCAATCTG 6480

TCATGAGACC TCCTGGCCTT CCTAAATCAA AGAGCAATGC CTGGGTCACG GAAGGGATGT 6540

```
ATGCCTTCTG TTACCGAGTG ATGTTTGAAG CCGGCTATCT AACACTGTTT GGCAGAGATA   6600
TTTCAAAGAC AGACACACAA AAAGCACTTA TTCTAAACAA CCTTGACAAC TTCAAACAAT   6660
TTGACCAAGT CTTTCCGGCA CTGGTGGCAG GCCTTCCTAT TCACTTGTTC AAGACCGCAC   6720
ATAAAGCTCG GGAAAAGCTG GCTGAGGGAT TGAAGCACAA GAACCTGTGT GTGAGGGACC   6780
AGGTCTCTGA ACTGATCCGT CTACGTATGT TTCTCAATGA CACGCTCTCC ACCTTTGACG   6840
ACATGGAGAA GGCCAAGACG CACCTCGCTA TCCTCTGGGC ATCTCAAGCA AACACCATTC   6900
CTGCAACCTT TTGGAGCTTA TTTCAAATGA TCAGGTAACT TTCCAGTGAC AGAAATTGCA   6960
TTTTAAACTC AAAACCCAAA AAGACTTATA GAGCTTTCTG TGCTATCAAC AAAGAAAGTA   7020
ATACTCAATG TCCGTGTTTA GCATGTGCGT AACAGAAGCA GCAATTTTTA GGTGCACAGT   7080
CCCATCGAAA GGGATGTCCC AGAAGCCACA GAACTCAGAC AGGTTGGTGC TCCATTAGTA   7140
CAGGTTCCCT GGCCTAGTCT TGCTCCTCAC CCGATATGTT CCTCTTAATA TCAAATTAAA   7200
TCCCCGAGTG CAGTCGTCAC CACCATATAA ACATTTGAAA TGATGACTGA CTTGCAGGTG   7260
TGATAAGAGC AGTGACCATA CCTTACTAAT TCACTGGAAT TCATAGGCAA AGTAACACCA   7320
TCGATTTTGT ATTCATATAG GAGCTGCAGC CATATTTTAA ATAGCACAAC TACTTGTTAG   7380
TCAAGCATTC TGAGGCTCAC TGTAATCAGG TAAAGTAGGT TTAACTCAGC GTCCTACCAG   7440
TTCCAGGCAT TGAAATGGAA TATCCTTTAT CCCACCCATT CAAAACGTAA TATATAAATG   7500
GAAGGCACAG TTTTGAAGGC CATGGTATGA TTTAGGGAAT TTACTCTCAT GGTCCAATCC   7560
CTTGTAATTG TATGCTAGGT GACATATCCT TCTGACTTAC TATGTTCATC GTATATTCAA   7620
TCCTTAGTTT ATAGAGACTG ACCAAAGCTC TGCTTTTGCA TAGCAAAGCT CCTTTTAATG   7680
CCCATTCCTA AACTCAAGGA CACGAATCCA GTTCAGTGCC CTTTTGCATA CTCCCTGGCA   7740
GACTCCCGTT GCCATACATC CTCCCTCGCT CGATTCCCAT GACCTCGCCC TTGCACACCC   7800
TGGTACTAGG ACCTCTCCTG GCGATACTTC CTACTACCTA TGCCACCTCA TTAAAAGGAA   7860
GGGATAATTG CTATTTACTT GCAGTTCTCT GAATGAGGAC ATTTTCCCCA TACGGCTCTT   7920
TCCACAGGAG TCCTGAAGCA ATGAAAGCAG CCTCTGAAGA AGTGAGTGGA GCTTTACAGA   7980
GTGCTGGCCA AGAGCTC                                                  7997
```

It was shown previously that high cholesterol diet up-regulates transcription of the cholesterol 7α-hydroxylase gene, translation of CYP7 mRNA, and increases enzyme expression and activity in rat liver (Li, et al. J. Biol. Chem. 265, 12012-12019, 1990). It is especially noteworthy that steroid regulatory elements (SREs) similar to those found in the LDL receptor, HMG-CoA reductase, and HMG CoA synthase genes are located in the upstream region of the rat CYP7 gene promoter. These SREs are not present in the human or hamster CYP7 gene promoter. These SRE's are

-1222-ATCCTCTCCCCAC TCCCAAGCATCCCTCCATG -1191, -1151-CAACTCCTCCCCTATT-1335.

Repeats 1 and 2 in the rat CYP7 gene are similar to the consensus SRE1 (CACC(C/G)(C/T)AC), which represses gene expression in the presence of oxysterols. The repeat 3 of the LDL receptor SRE has 11 bases identical to the sequence between -1151 to -1335 of the rat CYP7 gene. This sequence has been demonstrated to bind Sp1 which is a positive transcription factor in the LDL receptor gene (Dawson, et al. J.

Biol. Chem. 263, 3372-3379, 1988).

1(B) The Human Gene

A human genomic library, which had been constructed with Sau3A1 partially digested human placental DNA ligated into a BamHI site of the EMBL-3 Sp6/T7 phage vector (Clontech, Palo Alto, CA) was screened using a 1.6 kb EcoRI-PstI fragment of a human cholesterol 7α-hydroxylase cDNA isolated previously as a hybridization probe. Human CYP7 cDNA was isolated previously by Karam and Chiang, BBRC 185:588 (1992). Hybridizations were carried out at a high stringency condition of 68°C, 1% SDS and 0.1x SSC. 800,000 pfu of phages were screened. After four cycles of screening, seven positive clones were plaque-purified. Three clones comprising the largest inserts (λHGα26, λHGα5 and λHGα52) were isolated and analyzed by restriction mapping. Figure 2A shows the complete gene map of human CYP7. Clone λHGα26 (Figure 2B) contains a 15 kb insert which spans about 8.0 kb of the 5'-upstream flanking sequence and exons I to III (Tables 6 and 7) Clone λHGα5 (Figure 2C) contains sequences from intron IV, exons V and VI to an 8.0 kb 3'-flanking sequence (Table 8).

Table 6

```
TTTTTGGTTA TCTTTTCAGC CGTGCCCCAC TCTACTGGTA CCAGTTTACT GTATTAGTCG      60
ATTTTCATGC TGCTGATAAA GACATACCTG AAACTGGACA ATTTACAAAA GAAAGAGGTT     120
TATTGGACTT ACAATTCTAC ATCACTTGGG AGGCCTCACA ATCATGATGG AAGGAGAAAG     180
GCACATCTCA CATGGCAGCA GACAAGAAAA GAGCTTGTGC AGGGAAACTC CTCTTTTTAA     240
AACCATCAGA TCTCATGAAA TTTATTCATT ATCATGACAA TAGCACAGGA AAGAACTGCA     300
CCCATAATTC AGTCACCTCC TACCAGGTTC CTCCCACAAC ACGTGAGAAT TCAAGATGAG     360
ATTTGGATGG GGACACAGCC AAACCATGTC ACACTACCAT GCCTGACTTC CTTTCCATTT     420
TTGTATATTT GCTTGTTCTT CATTTGCCCG AGAAGTAACT CTAAAGGGCT GTATTATTTG     480
GATATTAGAT TGGCATTTTA TCTGACTGGG ATATCTTGCT GTGATTGTCC ATGTATAAGA     540
TCAGCTTTTC TATAAGCCAT ATTTTTAAAA AGATATATTA ATTTTTTAAA AATCCACCTG     600
TCTAAATAAA TGCACAAAGC CCCCCAAAAA CCTAGATTCT AAGAAAAATC TATGTACTGC     660
CATACAATGA TTGATATTAA TATTTATGGT GATAAATTAC ACACAAAAAA TGTGTGATCT     720
CTGTTTAAAC AGGCAAAAAC AAAAAACACA TGAAATAAAT CTATGGCATC TATAGCCAAA     780
ACTGGAAACA ACCCACATAT CCATCAATAG GAAATCAGTT AAATAAATTA TAGTACATTT     840
ATCCAATGGA AGATTAAGCA CATATTCAAT ATAATTATTT ATACACACAT ATAGATACAC     900
ACATGTATAA ATATAGAGAA TACTGTGGGT GTATGTGTGT GTGTGTTTAT ATACATATAT     960
ATACACACAC AGTACTGTTG CCTACCTTCT TTTGTCTTAA TTCTGTGAAC TCTCATTCAC    1020
TCTGCTTCAG TAGGATACCT CCTTCTTTTT GGTTCTTAGA CTCACCAAGT TGATCCTTGA    1080
CTCAAGACAT TGCATTTGCT GCTTCCTCTT CCTGGAATAT CCTTCCTTCT GATATTCACA    1140
TGAGTAGTCT CTTCTTGTCA TTCAGATCTC AAATGTCACA ATTTCAGAGA GCCCATCTCT    1200
GATCATCATA TCTAAAGTTG TCCTCATTCC CCCATAGCTT TCTATACCAT GTTTTATTTT    1260
TTTCATAACA TGTATTTTAT TACTCCTTTC TCCATTGGAA TAGAATCTCC ATTAGATTAG    1320
GAAATCTGCC TATCTTATTA ATGCCTGCAA CTGGAATACT TTTGAAGAGT TCTTGGCACG    1380
TAATAAATAC TCAACTAATA TTTTTGTGTA CACAGAAATA AAGTTTGGAA GAACAGATGC    1440
CAAATTGTTA CTAGTGGTTA CTTCTGAGTA AAGGAGTAGC ATGGTAGGTA AATTATTAAT    1500
AGATGTTCAC TTTCCACCAA GATATGTTTT AGTTAGTCTT AACTTACTTG AAATGAAATT    1560
TATTACTTTA ATAATTAGAA ACATTGATAA ACATTTTAGT CACAAGAATG ATAGATAAAA    1620
TTTTGATGCT TCCAATAAGT TATATTTATC TAGAGGATGC ACTTATGTAG AATACTCTCT    1680
TGAGGATGTT AGGTGAGTAA CATGTTACTA TATGTAGTAA AATATCTATG ATTTTATAAA    1740
```

AGCACTGAAA CATGAAGCAG CAGAAATGTT TTTCCCAGTT CTCTTTCCTC TGAACTTGAT 1800

CACCGTCTCT CTGGCAAAGC ACCTAAATTA ATTCTTCTTT AAAAGTTAAC AAGACCAAAT 1860

TATAAGCTTG ATGAATAACT CATTCTTATC TTTCTTTAAA TGATTATAGT TTATGTATTT 1920

ATTAGCTATG CCCATCTTAA ACAGGTTTAT TTGTTCTTTT TACACATACC AAACTCTTAA 1980

TATTAGCTGT TGTCCCCAGG TCCGAATGTT AAGTCAACAT ATATTTGAGA GACCTTCAAC 2040

TTATCAAGTA TTGCAGGTCT CTGATTGCTT TGGAACCACT TCTGATACCT GTGGACTTAG 2100

TTCAAGGCCA GTTACTACCA CTTTTTTTTT TCTAATAGAA TGAACAAATG GCTAATTGTT 2160

TGCTTTGTCA ACCAAGCTCA AGTTAATGGA TCTGGATACT ATGTATATAA AAAGCCTAGC 2220

TTGAGTCTCT TTTCAGTGGC ATCCTTCCCT TTCTAATCAG AGATTTTCTT CCTCAGAGAT 2280

TTTGGCCTAG ATTTGCAAAA TGATGACCAC ATCTTTGATT TGGGGGATTG CTATAGCAGC 2340

ATGCTGTTGT CTATGGCTTA TTCTTGGAAT TAGGAGAAGG TAAGTAATGT TTTATCTTTA 2400

AATTGCTCTT TGATTCATCC ATTTAATTTT TTTACCTTCA TTTTTATACA GTAAATTTGG 2460

TTTTCTATAC TTACACATAT TAGCATTATC TTCCTTATGT TTTAAATGAA AAATTTGATT 2520

TGAATTTTTA AAGTAATATC TTTTTTACTA TATCTCACAA GACATATGAC AGCTTCCCTT 2580

TTTAGTATTG GCATATACCG ATGGTAATAT ATAAATGTAT ATTGGTGTTA AACATAACTG 2640

ACAGAAATTG TATAAGGTCT CTATGTACAT TTATATGTGT ATCTAAAGAG GAAGCCCAGA 2700

TTAGTAAGGA TACAAGTAGC AAGTGGGAAT CTACAATGGA AAGGATTGCT TTCTCTCACA 2760

TGGCTTCAAT AGATACTCTT GCTTAAATAA ATGTTCTCTT TTAAGCTCAT TCTTGTGCAT 2820

CGCATAGACT CAGCCTAAGC CTGAACAAGA GCATAGAGCC TGAGCTGATC ATTCTATTAC 2880

TGTTTTTAAA TAAATGTTAA TCAACTGTGG TGAATTGGGA AAGTTTGCTG AGTGTATGTG 2940

ACATCGATTT CATTTATTTA CAACTGGTTC AAGAATGCAA GAAAAACAAA TACAGTCAGA 3000

TCCAGAACCA TAGTTTATTT AACTTCTAAT TGGCTCAAGG AGTAATTGTG GGGAGGCATA 3060

TAGATATTCT CTGCTATGTC AATCTCAAAA AGAGAAAATA ACCCTAACCA TCTTTCAGCT 3120

TTGTAGATTG CTATGTGTTT TCTGCCTTTG CAGTTTCTTT CAGGCCTGAT AGTTTTTACT 3180

TTTAATTAAA CTACTTATCT TCAAACTAAG AAAAGAAAGG TAATTACTTT ATACTGTATT 3240

ATTCTATCAA GAGGTACAGA AGTTTATGTT GGAAAATAAG TTTACATGTT CTAATAAAAA 3300

CATTTTAAAG GAGCACTGAA TTACAATAGA TGATTCCGTC AGTGTTTATC TTACTCAATT 3360

TCATTTTATA ATAAGCTGAT TTCTCACATG AGATTCTTCT TCTCTGAAAC CATCCTTATA 3420

GAATATAATA TAGATATCTT TAAACTAGGA ATATTTTCAA AACCTCAGTT CTGAAATCCT 3480

CCCTTATTCA GTGATCTGTG TCTTTAAAGA AAATAATCAA AAGAAACATT TTGAGATATT 3540

TAGAAAAATG ATGCTTAGCA AAGTGATAAA CACTAGAATG TAGTTTTGTT TCCGCACTGA 3600

CAACAAGAAT CTTGTTGGTC TTGTAAATCC TTTTGCCTGT ATCACTGGGA AAAGTGATGA 3660

GCACATAGTA GACGGGTGCT TGTTGAATGT GTATATGGAC GGATGCATGA ATGGATGGAT 3720

TTAGTAATCC TTTCCACCAA CATATCATGT TACTAGGTTA ATATAACCTA TTACTGTAGT 3780

AAAAGAGCAG GGCCCATCCA ACAAAAGAAA TATCTATAAA CTATAGGGTT TCAAAGTTTG 3840

AAGTCAGTGG GAAAAATTTT AAAACCTGAT GTAAGTAAAA ACCCAAAACT GTAATCATCC 3900

ATGTCTATCA TACACTTGTG TCTGACAGGC AAACGGGTGA ACCACCTCTA GAGAATGGAT 3960

TAATTCCATA CCTGGGCTGT GCTCTGCAAT TTGGTGCCAA TCCTCTTGAG TTCCTCAGAG 4020

CAAATCAAAG GAAACATGGT CATGTTTTTA CCTGCAAACT AATGGGAAAA TATGTCCATT 4080

TCATCACAAA TCCCTTGTCA TACCATAAGG TGTTGTGCCA CGGAAAATAT TTTGATTGGA 4140

AAAAATTTCA CTTTGCTACT TCTGCGAAGG TAAGCAGTTT TACATTTATA TACCATTCTG 4200

TTTGTCTTCT ACCTTTTTAT GTGCTTGTCT ATTTAGAAAT TTTGATGTAC TTAGATTTTA 4260

TGATAAAGGT GTTGAAGAGA GTTATCCTTA TGTGGAGATT CTTAGAAACA TAAATAAATT 4320

ATACGTAGCT TCTTAGTAAT AATCATTTAG AAAGTCAAAA TAGGTATAGA TTTCCGTCAT 4380

TTGCTTTGCA CGAGCTAATG AGGGTGAAAT ACAGATTAAA TGCTCTACTG AGACAGGTGG 4440

CACTGTACGA ATAAGATAGA TTAAAATTCA TCACATCAGC AATGTCTATG CAGAGCGAAG 4500

TGACGGAAAC CTAACATTCA GCAGTTGTCT CACCACACTT GTGCCACACA GTGTTTCATT 4560

TTGATAAGGA ATTGGCAAGA TATTTTAACA TCATTTAGAT GTAATAAAAG AAGATCTGTT 4620

ACTGAGAAAA AAAACCAATA ACTACTTACT TACTGCAAAT AAATATTAGC TTTGGTCTTT 4680

GTGACTAAGT AGCTTAAAGT TTGGTTAAAA TACATCTACA GCTGGACACA ATGGAACACA 4740

CCTGTAGTCC CTGCTATTTG AGAGGCTGAG GCAGGAGGAT CGCTTGAGTC CAGGAGTTTG 4800

AGGCTGCAGT GAGCTATCAT TGTGTCACTG CACTCCAGCC TGGGTGACAA TGTGAGACCC 4860

CATCTCTAAA AGAAAAAGAA AAAGAAATCT ACAAATAATA TAAAAGATAA CTAATGATTT 4920

TAAAACATTA TCAATTAGTT TATGTGCAAT AGCTGTAAAT AAGTGCAGTA GCATAAGAAA 4980

TAAGACATAG ATGACTTGAG TGATCCAGGG GAGTGCCACT GAAGTTGGCT TTAAAGGAAA 5040

GGTACAGTTT GGTCATTTAT TTGTAAAGTG CTATGAACTT GTACAAGGGA AAGCCAATTT 5100

CCCGTGTTTA CCAAGTAAGG AACTATGAAA GTATCTAATC CGTTTTTCAG TCATTTACTA 5160

TGACTAGGTC AGGTTTAACT TCTTTTTCTG CATGTTTTAT TTGCTATCAG GCATTTGGGC 5220

ACAGAAGCAT TGACCCGATG GATGGAAATA CCACTGAAAA CATAAACGAC ACTTTCATCA 5280

AAACCCTGCA GGGCCATGCC TTGAATTCCC TCACGGAAAG CATGATGGAA AACCTCCAAC 5340

GTATCATGAG ACCTCCAGTC TCCTCTAACT CAAAGACCGC TGCCTGGGTG ACAGAAGGGA 5400

TGTATTCTTT CTGCTACCGA GTGATGTTTG AAGCTGGGTA TTTAACTATC TTTGGCAGAG 5460

ATCTTACAAG GCGGGACACA CAGAAAGCAC ATATTCTAAA CAATCTTGAC AACTTCAAGC 5520

AATTCGACAA AGTCTTT 5537

Table 7

GAATTCTACT CTTTAAAGGG GTGAATATTA TGGTACTTGA ATTTTATCTC AAGAAAAATG 60

AATAAAAAGT AACTAAATCA TTGAAAATAT CTGATGGCAT GGGGTTTGTG GGGTAACTGG 120

CATTCCACAG TGATTTTCAA AGGGCTTGTG CTGTTTTCAT TTTGCTTTGT TTTAGTTATG 180

GAGCCCTTCC TTGAAACAAA CTTCATACTA CAGTCCTCTT TCATGAAGCA GAAGAGGGCA 240

GTGGGCAGAG CTCTCCTTTG GCTTTCTCCC CCACCACAAC AGGGAGCCCT GGAGCTCTAG 300

GAGAGAAAAT CTGAAATATA AAGGGCATGC ATGTGAGCTG TGGAGTCCCA GAGCCCTGGG 360

TTTGCATCCT AGATCTGCAA CTCCCGTGAA TTGAGTTTTG GGAAGTTGCT GAAACTCTGA 420

CCTCCTGTTT TCTCATGGTA TTGTTGTAAG GGTTAAATGA GACAATGTAT GTGAAGACCC 480

TGGCCCCACA GTAGAGGCTC TGCACACATT TCAGCGATAC TTTCCTCATG TATTTCCAAA 540

AATGTTTTCT CATTTTCTTA AAATGTCAGA AAGAAGACAA CAGAACTTAC TTGCCTTTTA 600

CAACAGAACA AATGGAGCAA GTCAGAGGTC AAGGTGCTAA CATTCTTCAT GGTTCCTCAC 660

CACCTTTTGT TCTGTTAGCC TATAGGGAAA AGTCTTCTTT CTCATCTCAT TATCTGCAGG 720

GGAAAATAGT ACTTCAGCAA GTGATCCAGT TGAAGAACAT CTCCAGGGCC ATTAACATAC 780

AGAGGTTTGT TCTACTCTCT CTGTGCTCCA TGTCTAAGAA CCTCAGCCTT CCTCCTAGGA 840

GCTAGGGAAA GTCAGGAAAG TGAAAATAGT ACCCCAGCTA ATGAACTGCC CTGTGCTGGC 900

CTGAGAAGAC AAGACCAGCT TCCTCAATGG CTCAAGATTT GGTTTCCTTC AATATGTCCT 960

TTTGGAAATA TGTCCATGAC ATCGGAGAGA TAAAAGGAGC CAGGATTGCT CACATTCAGG 1020

AAAAAAGCTC CACTATCTTT CTCTCTCTCC CTCTTTCTCT CCCTCCCCCT GACTGCCCTC 1080

TTCTCTATCT CTCTCTCTCC CTGAGCTGGC AAGGTTAATT GGTCGCAGAA AGCCGAAGAA 1140

ACAAGTGGGC CTCCTGGAAC AAAGTTCAAA AAGCCGAAAA CGGGAAGAAA ACTAACCACA 1200

AAAGTAAAGG AACCACTTAG CCTTCTTTGA TTCCAGGCCC CCAAGCCTGT CTTTAACTTG 1260

GATGAATGGA GTTCTTCCTG TGCTACAGCA CCGCATAGTA GGGGCTGCCC TGGGCCTGAA 1320

GCCAGAGCTT CACCATATTC AGTCATCTGT ACATTGAGGC AACAGTGCCT GCTTCATGGT 1380

GCTACCCTGT GGATTAAATG AAGCAAGTTT TTGATGATCT TGACACTGAA TATTGATGCA 1440

TTGGTCAGAC TTTTTCTGAT AGTAAAAAAT GGTGGTTTCT TGTTGTCAGA AATCAAATCA 1500

ATATATTTGT TCTCCTGTTG ATTAGCTATG TCCCCTAGAG GGCAGCGACT TTGCCTGTCT 1560

TATTTATCTC TGCATCTCCA GCACTTAAAA GGTGCCTTGC ATAAGGTACA TATTAAGTTC 1620

ATATGAATGA ATGAATGAAA TGCATATGAT TTATTCATAC CCAGTTGGTG GTGTGTTTAC 1680

CCTTTCCTAA ACCTGTAGTC AGATGGCCTT TGAATCCCCT GTACTTCTTG TGAGGTACTG 1740

TGCTGTAAAG GTGGACTATC ACACTTCAGT TCAGAGCAAT CTGGGCTTGA ATCCTGGATT 1800

TGCCAGTTTA TTAACTATAG CAAACATTTT TGAGCATACA TTGTGCCAAG TGCTAGGCTA 1860

ACTGTCTTAC ACACATTGTC TTATTTCGTC TTAATATCTA TGAGTCATGC ACTATAATCA 1920

TCCCCATTTT ACAGATAAGA AAGCAAAGAC TTGGAGAGGA AAAGCATCTT GTTCAAAGGT 1980

AAATACTTAA TGGCCAAGCC AACATGCAAA TCTAGATTTA ATTGCAGCTT CCTCTTCATC 2040

TACCATTCGA ACTAATTCAA GCTATGTAAT ATTTCCCACT GAACCTTCTT GCCTCTACTT 2100

CCTCATCTTT AACATGGTCA AAATACCTGT CCTGCCCAAG TTAGTTATTT CATTAAAGTA 2160

GAAAAATACA AGAGAAGCTT TTAAAATGTG AAACCTCAAA TGAATGTAAA ATTATGATGA 2220

TTCCTTTAGA ATTTGTCAAC ACCTTCTTTT CTCTACTCCT GCTAGGCATT TACAATCTCA 2280

AAACCATGTA TTTAAGATGC AAAACTATAT TTGTATTTGC CATAACTGGT TTCTTTCCCT 2340

ATGGCTTCAT GAAAATGTGG CTCGAATGTG TTTATTATGA AAGCCCCAAA TTAATCACGA 2400

CAAGACTTCA CCAGCCCATT CCACAATAGA CTCCCATTAC TTTGCCCTGA CTTAGAAACC 2460

TCATATACAG TCTTGATTCA GTACAGCTCT GTGATGCTCT TGGAAAATGC AAAGTGCTTT 2520

CTTAATTGAG GCAATCTGTG TCCCACTACA GAGAGGTGGT TTAACTTGTG AATTC 2575

Table 8

AGAGCAACCT GGGCAACATA GCAAAACCCT GTCTCTGCAA ACAATAAAAA GAAGAAAATT 60

AGCTGGGTAT GGTGGCACAT GCTATAGTCG CAGCTACTCG AGAGGTTGAG GTGGGAGGAT 120

CAGTTCAGCC TGGGAGGTTG AGGCTGCAGT GAGCCAGATC ATGCCACTGC ACTGCAGCAT 180

GGGCAACAGA ATGAGACCCT GGCTAAAAGA AAACAAAATA AAAAATTCAG ACACAGGTTG 240

AATCATTGAT AACAGCATAG TGGTAACAGA AAGAAAGTTT GGGAAATTTT TATCTGATCA 300

GCTTCCCATA CCCTGTTCAT CTTTGTGTTA TGCACTGCCA GGCTGTCTGT AGGTTCAGAC 360

TCTATATCAT ATGACCTTCA AACACTTGGT TTGTTCTTCT CCTTCCTTCC TCCCTTCTTC 420

TTTCATTTTT TATCTTTTTT TCTTTTAAAA TGTTTAGATA GTATAATAAG GAACTGCTGA 480

GGCTTTCCAG TGCCTCCCTC AACATCCGGA CAGCTAAGGA GGATTTCACT TTGCACCTTG 540

AGGACGGTTC CTACAACATC CGAAAAGATG ACATCATAGC TCTTTACCCA CAGTTAATGC 600

ACTTAGATCC AGAAATCTAC CCAGACCCTT TGGTAAAGTC GCAGTGTGCC CGAATTGAAA 660

```
TTCAATATCC AGGTGATAGC TACCTAGATC TAAATAAAGA GGAAATTTAC AATGGTAGAA     720
TTGATTTTCT CATAGTAGTC ACAGGAATTG TCTGACTTAA TTGTGTTAAA TATTCATATA     780
TTTTGGAAAA TTTAGATAGT GGTCTGAATT TTTCATTTTA GTCCTGATAT TTGCCATCAC     840
ACAGTCTTTG CTAGATTATA TTTGCAGTCA TGATAATAAA CCTGCCACTT TTTTTTCTT      900
AAAAAGCACC TCCTCCCAAA TCCAGGAAAT TGGAGGCTAA TATATTGATT ATTCTAGTTT     960
CTTCTGGGAA CCCTTCTCTC TCTAGCTCTG CCTGACTAAG GAACTAATCG TTCAAGCAGG    1020
ATAGGAAGGT ATCACAAGGC TTCCTTAGCT GCATTAAGCT CCTGTTCCTT ATTACTTTCT    1080
GATTCAATGT GGAGTATTTG CTAAATCACT AATGGGGTAG AATTAAAAAG AAAATTACTC    1140
TTTGGAGCTT CCAGGTTTAG AAAGAGATAA ATTTCTTTAA AACTAGCTTA AAGGCGGTTT    1200
TCTTTGTATT TTTATTGCAG ACTTTTAAAT ATGATAGGTA TCTTGATGAA AACGGGAAGA    1260
CAAAGACTAC CTTCTATTGT AATGGACTCA AGTTAAAGTA TTACTACATG CCCTTTGGAT    1320
CGGGAGCTAC AATATGTCCT GGAAGATTGT TCGCTATCCA CGAAATCAAG CAATTTTTGA    1380
TTCTGATGCT TTCTTATTTT GAATTGGAGC TTATAGAGGG CCAAGCTAAA TGTCCACCTT    1440
TGGACCAGTC CCGGGCAGGC TTGGGCATTT TGCCGCCATT GAATGATATT GAATTTAAAT    1500
ATAAATTCAA GCATTTGTGA ATACATGGCT GGAATAAGAG GACACTAGAT ATTACAGGAC    1560
TGCAGAACAC CCTCACCACA CAGTCCCTTT GGACAAATGC ATTTAGTGGT GGCACCACAC    1620
AGTCCCTTTG GACAAATGCA TTTAGTGGTG GTAGAAATGA TTCACCAGGT CCAATGTTGT    1680
TCACCAGTGC TTGCTTGTGA AATCTTAACA TTTTGGTGAC AGTTTCCAGA TGCTATCACA    1740
GACTCTGCTA GTGAAAAGAA CTAGTTTCTA GGAGCACAAT AATTTGTTTT CATTTGTATA    1800
AGTCCATGAA TGTTCATATA GCCAGGGATT GAAGTTTATT ATTTTCAAAG GAAAACACCT    1860
TTATTTTATT TTTTTTCAAA ATGAAGATAC ACATTACAGC CAGGTGTGGT AGCAGGCACC    1920
TGTAGTCTTA GCTACTCGAG AGGCCAAAGA AGGAGGATGC TTGAGCCCAG GAGTTCAAGA    1980
CCAGCCTGGA CAGCTTAGTG AGATCCCGTC TCCAAAGAAA AGATATGTAT TCTAATTGGC    2040
AGATTGTTTT TTCCTAAGGA AACTGCTTTA TTTTTATAAA ACTGCCTGAC AATTATGAAA    2100
AAATGTTCAA ATTCACGTTC TAGTGAAACT GCATTATTTG TTGACTAGAT GGTGGGGTTC    2160
TTCGGGTGTG ATCATATATC ATAAAGGATA TTTCAAATGT TATGATTAGT TATGTCTTTT    2220
AATAAAAAGG AAATATTTTT CAACTTCTTC TATATCCAAA ATTCAGGGCT TTAAACATGA    2280
TTATCTTGAT TTCCCAAAAA CACTAAAGGT GGTTTT                              2316
```

Cloned bacteriophage λHG7α26 and λHG7α5 were deposited August 25, 1993 at the American Type Culture Collection, ATCC, 12301 Parkland Drive, Rockville, Maryland 20852, U.S.A., under accession numbers ATCC 75534 and 75535, respectively.

Five EcoRI fragments of the clone λHGα26 were excised from the phage DNA insert by restriction digestion and shotgun subcloned into the phagemid vector pBluescript II KS + (Stratagene, La Jolla, CA). The clones were size-selected. EcoRI fragments were isolated from CsCl purified plasmids and used for sequencing. Nested deletions were generated by ExoIII/Mung Bean nuclease digestion according to the manufacturer's instruction (Stratagene, CA) using the conditions of a 37°C incubation for 1 min intervals. This condition resulted in an average deletion of about 200 to 250 bp/min. DNA sequencing of the nested deletions was carried out by the dideoxy chain termination method using T7 sequence version 2.0 (USB, Cleveland, OH) and ^{35}S-dATP. Sequence data were obtained from both strands and the overlapping

deletion clones and analyzed using DNASIS software (Hitachi America, CA).

The nucleotide sequences of a 5.5 kb EcoRI fragment (Table 6) and a 2.6 kb EcoRI fragment (Table 7) were determined. The 5 kb fragment contains the sequence from -1886 of the 5'-upstream region to a partial exon 3 (Figure 2B). Included in Table 6 also is the 347 bp 3'-end sequence of a 3.5 Kb EcoRI fragment located immediately upstream of this 5.5 kb fragment (Figure 2B). As shown in Fig. 2A, the 2.6 kb fragment is located further 5' upstream of the 3.5 kb EcoRI fragment. Thus, a 4823 bp 5'-upstream flanking region sequence of the gene now is determined.

Molowa et al. (Biochem. 31, 2539-2544, 1992) published a 1.7 kb upstream sequence of a human gene. A comparison of the sequence of the present invention to that of Molowa et al. in the overlapping region (1604 bp) revealed that sequences from the transcription start site to about -460 are identical, however, further upstream the sequence vary significantly. A total of 52 sequence discrepancies were found, which are far too many to attribute only to the presence of polymorphisms in the human gene. Cohen et al. (Genomics, 14, 153-161, 1992) reported a 723 bp upstream sequence and suggested sequencing errors by Molowa et al.. Thus, the sequence of the present invention, from the transcription start site (nt + 1) to -587, is identical to those reported previously by Molowa et al., Nishimoto et al., (Biochem. Biophys. Acta, 1122, 147-150, 1993) and Thompson et al., (Biochim. Biophys. Acta. 1168, 239-242, 1993).

The present invention identifies seven mismatches in Cohen's sequence from + 1 to - 123. A conversion of at T to C nucleotide -469 was identified to be a Mae II polymorphism (Thompson et al., 1993). The 5'-flanking sequence of the present invention agrees very well with that reported by Thompson et al. (1993). Only one mismatch at nucleotide -1193 (C vs A) was found in the overlapping region from + 1 to nucleotide -2235.

The present invention further identifies transcription factor binding motifs in the human gene, however, SRE-like sequences were not found in the human promoter region.

1.(C) The Hamster Gene

A hamster liver genomic library constructed in the λDASH II vector (Stratagene) was screened with a 2.5 kb Eco RI fragment of the rat pBSK7α12 comprising the entire coding sequence of the rat cholesterol 7α-hydroxylase cDNA. About 1 million plaque-forming units were screened and one positive clone was identified and plaque-purified. The phage DNA was purified by CsCl gradient centrifugation and cDNA insert was restriction-mapped using rat probes (Figure 3). EcoRI fragments of the DNA were isolated and subcloned into a pBluescript II KS + vector. Nested deletions were generated with an ExoIII/Mung Bean deletion kit. The DNA sequences of these deletions were determined by the dideoxy chain termination method using Sequenase. In some instances 17-mer synthetic oligonucleotides were designed and used as sequencing primers. Sequences were determined on both strands with overlaps. cDNA sequence analyses were carried out with DNASIS software.

Table 9 shows the 11 kb DNA sequence of the hamster gene. It covers the sequence from nucleotide-1650 of the 5'-flanking region through all six exons and five introns (Exon I: nucleotide 1651-1730; Exon II: 3511-3650; Exon III: 4351-4937; Exon IV: 5945-6075; Exon V: 7690-7865; Exon VI: 8437-8736). The amino acid codons interrupted by introns are identical in each of these three homologous genes. The DNA sequence of the exon-intron junctions follows the canonical GT-AG rule typical of eukaryotic genes. The precise intron sizes determined by DNA sequencing are consistent with those of the rat. The intron 3 of the hamster gene is 1007 bp, which is about 1 kb shorter than that estimated for human intron 3. A putative polyadenylation signal (AATAAA) is located 371 bp upstream from the 3'-end of the gene, indicating that the isolated genomic clone should include the entire coding exon 6.

Table 9

```
GAATTCTAAA CACATATTAA TATCAATGAC TTATATGTAT GTATATATAT ATCTAATATA    60
GATAATGTAT CTAGGGATAT ATATATATGT ATATTTTATC TTTCTTCCTT TTATTCTTTC   120
TTCTCCCCTC TCTGTTCAAC ACCGAGGAAT AGAATGCACT GTGGTGTCAT ACTCTGCTTA   180
CTCAGCCTCT TATTGACCTC TGAGTCAATA CAGTGCTGAT GTACATCTCC AAATGCCCTC   240
TTTTCTCCTA ACCACAGACT TTTACATTCA GTAATCAATT TGACATTGTC CCATGATTTA   300
CAAATGTTCA CAATAGTATA TTGACCTATT GCTGCCTTCC AAGGTCCTCT CCCACTCCCA   360
AACATCCCAA TATGAACCAG CTTTTGCCTA TCTTCTTGTC TCTTACTTTA ACTCAATGTC   420
ATTCCCTATT CACTTTGCTG TAATAGATGC TACCTTGATT CTGGTTTTTA GCACCTTAAT   480
TTCGCTCTCT GCTCAGGAAC TCTGCCTTTG CTGTTCCCTC TTCTGGGAAC GCTTTTCCTT   540
TGCTGTTATA TCTCTTCAAA ACAGCTTCTC TATTCAATAT GCTCAAGCTG CCTTCAGCCC   600
```

TCAACAGCTC TCCCTACCTC ATTCTAGTCC CTCCACTAGA ATAGAATCTT CATGAGAGTA 660

GCGAACTTCC CTATCTTGCT AGTACCCAAA GGCAGAAAAA TCTTTAAAGA GTTCCTGGGA 720

CATAGAAAAA GTGCTCAATT AATATTTGTA TTAAATAGGG ACCTCAGGTG TAACTCCGTG 780

GTAGAGCGTT TGCCTTAGAG AAGTAGGGCC ATGGGTTCAA ATTCCAGCAC AGAACAAAAA 840

ATTGTGCTGA ATAAAGTTTG GGAGGATGTG TAGCAGTTTA TAGTGCAAGT GGCATAAGCA 900

GTAAATAATG AATTTGTATC CACTTTTCTA GCAAGAAGTA TTTTATTCTT TATTTGAAGG 960

ATAACAATTG GTAAAGACTG CATTCTCAAA ATAAACTATG GCTTATGGCT ACGTGGAAGA 1020

TGAGATAGGG AGAAGGTTTT TTTTTGATGA TGGCAAAATA ACATGTCATA GTCCACACGA 1080

AACACCTGTG AAGTTGTAAA CACACCTAGC AATCAAACAA GAAAATTGTC CCACCCTATT 1140

ATCATTCTTT TGGATTGGTT GTGGCATATT TCTGGAAAAT GATTTAAATT AATTCCTTCT 1200

AAAGGTAACA ACACAAACAA CCACTATCAT GACGAAAAGC TTCTGCCTGT TTCAGTTTAC 1260

ATCATGCTCA ATGTCTACAA CAGACGTGCT CATCTTCAGA GTGTTTACCT CTGCTTTTTA 1320

CACACATTGA AGCACAATGT GAGCTGCTGT CCCTGGGTCT GAATGTTATG TCAGCACACA 1380

AGGGACAGAG CTTCGGCTTA TCAAGTATTG AAGCTCTCTG CTTGTTTTGG AGCCTCTTCT 1440

GATACTATGG ACTTAGTTCA AGGCTGGGCA ATACTATTTT TTTCTTTTTT CTAATAGGAG 1500

GACAAATAGT TAGTTGTTTG CTTTGGTCAT CCAAGTTCAA GTTATTGGAT CATGGTCCTA 1560

TGTGTATAAA GAGTCTAGTT TGAGCCTTTC AGGGGCAGCC TTGCTGGCTA AGCACAGACT 1620

CTCCTCTTGG GAGTTTTCCT GCTTTGCAAA ATGATGACCA TCTCTTTGAT TTGGGGGATT 1680

GCTATGGTAG TGTGCTGTTG TATATGGGTT ATCTTTGACA GAAGGAGAAG GTATGTCTTT 1740

TAGCTTATTT CTAGTGTTTT CACTATTATA CAGTTCCAAA AAAATACTAG TACATTAGTA 1800

TTTTTATTTA AAATTTAAAG CCATGCTTCT TTGACTAAAC CTGACAAGAT GTAGAGTTTC 1860

CCTTTGAATA TCCACATACA CTGATGGTAA TGCTGATCTT GTTAAACATA ACTAAAAAAA 1920

TTATAAGTAT TGATGCATGT TTGTGTGCAC TTCTGTGGAG TACACCTAAG CTGGGAAGGG 1980

TGCATTTGGC AAGGGTGACG TTTGGAAAGG ATCTTTCTCT CACAATAACT GGTTATGCAT 2040

ATGCTCTTCT GGGTTCTCTG TTACATCAAC ATTAAAATAC AGGAATACCC TTGGCATATC 2100

TTTGGCAAGG TAGACTGTGT CTGCTGTCTT AGTTTTAATA ACTTCTTTGC CTTTTGAGTT 2160

ATTTGAATTT ATGCCTGATC GTTTCCAGTT TTAGTTGTCT TAATGCTAAG AAAGGACAAA 2220

TCAATTATAT TTAGTTATTC TAACAAGAGA TAACTAGTTT ACGTTGAAAA ATAAATTATC 2280

TTATAATTTC TAATAAAAAC ATTTAAGAGA GTTAGAAATC AGCGAATTAT AGCTGATGAT 2340

CTGCCAATGT TTACCTCACT CAACTTCATT TTAGATACTT TTTCAAGTGG GATTCCTATT 2400

CTCTTCAAAT ATCCGCACAG AATTATAGTC CCCTTCTTTC AGAGTGGGGG GAATCAAATG 2460

AAAGGTTTCA TGTGTGCTAG GCAAGAGCAC CACCGTTGAG CCACACCTCC AGACCCCACA 2520

ATGCCAACAT TTTTAAACTA TGTAGAGTTT AAAAAACTTT AGTTCTGTAG CCTTTTCTAT 2580

TAGCTGGTGT TTCATGTCTT CAAAGAAAAG GAAAACTGAA ACATTTTAGA CATATGGACA 2640

AATGATTCCT TGAACAAGTC TAAGCACTGA TGATAGCTTC TTTTCTACAG TGAGATCAAG 2700

AATCTTGTTA GCCCTGTTGA TACTTGTAGC CCTGTCACTT GGAAAAGCAA TCAATTTTAT 2760

GATCTAGAAA ATAGAGCTTG CCTAAAGATC AGAGTGCAGA GCTAGTCACA CTAGTCAGCC 2820

ATACAGGTTA GGCAGTGGTG GCACATACCT TTAATCCCTG CAGCCACTCA AGTTACCCAT 2880

AGAAGCTGGG TGGTGGTGGT GCACACCCTT AATATAAGGT GGAGCACACT TTAATGTAAG 2940

GTGGGTAGAG TCAGGAGTGC AGTGTATTCA GTCTGCAGTC ACACTGAGAA CAATATCACC 3000

CCAGTCTTGT TAGAGGTAAG AACTCTCTAG TGATTGGCTG CTTTGCTCTT CTGATCTTCA 3060

GTTTGAACTT CTGTCTCTGG GTTTTTATTA TTCGTGCTGC AGACATAGAC ATAGCAAACA 3120

ATTTAATGAG TGATTGATGA ATGTAGATAT GTATGTACAT ATTGTGCTGG ATAGACTGTA 3180

GATGGGTTGG TGGATGGGTT GATGAGTGGG TAGATTTAGT AATCACCTTC ACCAATATCT 3240

TAGTAGGCTA AAAAGCCCAC TGTTTTAGTA AAAGAGTGGG GTATCCAACA AAGAAGTATC 3300

TATAAACTGT AGTTATGTGG TAGAAATAAG GGGTAGAAAC CAGTAAAAAT TCGGCTTATG 3360

TACAAATGCT AAACATGTAA TTTCCTAAAC CTCTCAATCT GTCTCACAGG AAAGCAGGTG 3420

AACCTCCTTT GGAGAATGGG TTGATTCCAT ACCTGGGCTG TGCTCTGAAA TTTGGCTCTA 3480

ATCCTCTTGA GTTCCTGAGA GCAAATCAAA GAAAGCACGG TCATGTTTTT ACCTGCAAAT 3540

TAATGGGGAA ATATGTTCAC TTCATCACAA ACTCCTTGTC ATACCATAAG GTGTTATGTC 3600

ATGGAAAATA CTTTGATTGG AAAAAATTTC ATTACACTAC TTCTGCAAAG GTAACTAGTT 3660

TTTACAGATT TTGCTTGTTT ACTAGCCTGT TTATTTATTA GTTTATTTAG TTGTTCCAAT 3720

GTTATTAGAT TGTAGGATAA AGGGAACATA AAATCAGGAA GTCTCTTGGT ACTAAGCATT 3780

AAAAAGTCAA GGTAAATGTG AATTTGTGAT TGATGATGAC ATACACAAAT TAAGCACTTT 3840

GTAAGTACTT TCTGAGCCAG AAGACACTAC AGGAAGGCAC AGACTCATAA CATCCATGCT 3900

GCCATCTACA CAACACTCAG AGCACTCAAT TACCACATCA TGCACACGAA CTCGTTCGTT 3960

AAGAAGTCGA CAGTATATTT AAGCATCATT CAGATGTTAT CAAGAATCTC TATTCTAGAG 4020

AAAACAACAC TTAGCTGAAT TTTTACAAGA AAATATTAGA CATGGTCTCT GTCTTAAGTA 4080

GATTAAAGTC TGGCTAAAGT GCATCTGCAG AGAACAAAAG GTAAAGATAA AATCAATGGC 4140

CCATTAGTCC AGAGAAGCTT ACCTGAAAAT CTGGGATTTA AACTTGACCT TAAAGGAAGA 4200

GTATGTCTTA AGTTTGACTT TGAAAAATGT TATGAAATTG TATTGGGAAG GCTAGACAGA 4260

GAAGTATGAT ATACTTTAAT CCATCTTCCA GCCATTTCCT AACACCCAGG TTTAGCTGCT 4320

CCCCCTCTGA CGAATTTCAT TTTCTACCAG GCATTTGGAC ACAGAAGCAT TGACCCAAAT 4380

GATGGAAATA CCACAGAAAA CATAAACAAC ACTTTTACCA AGACCCTCCA GGGAGATGCT 4440

TTGCATTCAC TCTCTGAAGC CATGATGCAA AACCTTCAAT TTGTTCTGAG GCCTCCTGAT 4500

CTTCCTAAAT CAAAGAGTGA TGCCTGGGTC ACCGAAGGGA TGTATGCCTT CTGCTACCGA 4560

GTGATGTTTG AAGCTGGATA TCTAACTCTG TTTGGCAGGG ATACTTCAAA GCCAGACACA 4620

CAAAGAGTGC TTATCCTGAA CAACCTTAAC AGCTTCAAGC AATTTGATCA AGTCTTTCCG 4680

GCGTTGGTGG CAGGCCTCCC TATTCACTTG TTCAAGGCGG CACATAAGGC CCGGGAACAG 4740

CTGGCTGAGG GCTTGAAGCA TGAGAACCTC TCTGTGAGGG ACCAGGTCTC GGAACTGATA 4800

CGTCTACGCA TGTTTCTCAA TGACACTCTC TCTACCTTTG ATGACATGGA GAAGGCCAAG 4860

ACACACCTCG CTATCCTCTG GGCCTCTCAG GCAAACACTA TTCCTGCAAC CTTCTGGAGC 4920

TTATTTCAAA TGATCAGGTG GATAGCAATT TGAGTGTTTA TTCTTCATAG TGACAGAAAT 4980

TAACAATTTT TAATAAACCC CCCAAAAGAC TAGCAGAGCT TTCTTTGCTG TTGGTCAAGA 5040

ATGTGATACT CAGTGCCTGT GTTTGACATA TATATATAAC AAAAGTAGCA TTTTGTAAGA 5100

ATATAGTCTC ACCAGAAAGG GATGTCCCAG AAGCCGCAGA ACTTAGATCT GCTGGCACTT 5160

GTCATTAAAG GTCCCCTTGC CCAGTCTTGC TTTTAACTCC ATAGTGTTCT TCTTAGTGTC 5220

AAGTTAAATC TATGACTGCA GTCTTCATCA CAACTTTAAA TAATGACTGA CTTGTCAATG 5280

TGGTAAGTGC AGAGGCCACA CCTTACTAGT TTGAACATTC CTGTTTTCTG CGGCCTCACA 5340

GATTTACAGC AGAGTTGCAA CATCAATTTC ATATTACCTA TGAACTACAA CCATATTTTA 5400

AGTTCAACAA CTACTTGTTA GTAACATTTC TGAGGCTCAG TTCACTTTAA CCAGATAAAG 5460

GAGATTTCAA ACAGCTGCCA ACAAATTTCC ATGCACTGAA TGGAAGTATT CTTTATCGCA 5520

CAGTTCAAAA ATAATAACAT AAATATTCTG AAGCTGTGGT ATGAATTTAA AGAGTAAATT 5580

TGAATTTCTA CTTGGGAATT CACCAATACC CTGTAATTGT ATGTTAGAGG AAGTATTCGG 5640

AATGAATTAC TCTACTCATC ACACGAATGT CTAGCCCTTA TTAGAATCAT TGGTTTATAG 5700

AGATCTGACC AAAGCTTTGC TTTTACATAG CAACGCCCCT TTAATGCTTC TTCATAAATT 5760

CAAGGACATG AATCCAGTTC AGAATACAGT ACAAGTAAAT GACAATGCCC TTTGCATGTT 5820

CCTGGAACCA CTTCCCTTTT CATGCTCCCA TGCTAACGCG ATCACCTCAT TAAAAGAAAT 5880

GGAGTTCTTA TTTACTTGCA GCTCTCTGAA TAAGGCAATA TCTTCCATAT GTCTCTTTTC 5940

ATAGGAGTCC TGACGCATTG AGAGCAGCCT CTGAAGAAGT GAATGGAGCA TTACAGAGTG 6000

CTGGTCAAAA GCTCAGCTCT GAAGGGAATG CAATTTATTT GGATCAAATA CAACTGAACA 6060

ACCTGCCAGT ACTAGGTGTG TTCCCTATGC TATCCCTCAC TAACATGTCA CTAGTAACAA 6120

TGCTCAACAT ATAATGAATG TACTATATTC TTGATATTTT TGCAACGCTG CAACAGTCTA 6180

ATAACTAGGG TCATCTTCAT TTTTTCTAAC AAACAAGGAA CTGAGACCCA GAGCGTGGGA 6240

CAGTGGCAAC CCTGGCATAG AACATTTGAT ACTCAGTTGC TCTAGGTCCT TGGCCTCCTT 6300

TCTTAGTCCT CCAAAACCAC AAACCCAGGG TTAAGGAAGC ATGGAATTAA TGTGAACAAA 6360

GCAACACCAT TGGTTTGGGC GATGAGACTG AGGCTTTTCT TCCTTTGTTT CTGTATTTTC 6420

TAGAATGCAG TAGTACCATG TATTACAGTA AAACAGCCAT ATTTTTGTGT CCTGTTCTGT 6480

AAAGGACAGA AGCCCCCATA TGCTTTGAGG GCAGTTTAGT TTATTAGAAG CAACAGAGCC 6540

TAGATTCAGC ACTGCCTGGT TTGGGACCTC CCTTTAGACA CCTCCCTTTT CTCACCTGTA 6600

AATAAAGGCT AAGTAAGCAT TTGTGACTGC ATACTCAGTC ATGGCCTGAA TCCTGGGAAC 6660

AAGGCAGCTA GCAGCTAGAG GCTGGAAAAC AGGACTGGAC CTCAGCAGCT CTACTGCATT 6720

ACTTCCCCTA GAAGCAGGGT GTGGCTACAC AAAACCAGAC AGATAATGTA TGGCTGAATG 6780

TAGATTCATG AAATGCTTGG AAAGACATTT ACTTATCAGT ATGTTTAATT CCCAAAATGG 6840

TCAGCAACAA TTCACACAAA ATTGATTATA AGTTTTTTCA ATTTGCTTAG CTGTTTAGTG 6900

TCCAGTAGAA ATAAGATTAC TATTCTATAA AGTGACAGAT GTTCATCTAG TTCCCATTGA 6960

TGGTGAAGAA CATTATGTCA TCCCAAAAGA TCGTTAACTT AGATCGTGGT TCTCTACCTT 7020

CCTGATGTTG TGTGACCCCC AACTGTGAAA TTATTTTCAT TGCTACTTCA CAACTATAAT 7080

TTTGCTTCTG TCATGAATCA TAAAGCAAAT ATCTGTGTTT TCTGATGGTC TTAGGTGACC 7140

CCTGTGAAAG GGTCATTTGA CTCTACCCCC TACATGGGTT GTGATCCACA GGTTGAGAAG 7200

CACTGACTTA GATTCTCAGA TTGCAAGTAG AGCAGCAGAA TTTCGAAGAA CAGCAGTGGC 7260

GACAGAAGCT GCTTTGGGCA GTTGTCATTT GTTAGCTTTC ATTGGCTCAT TTTGTATACA 7320

GATTTTCGGA AGTATTTCAG ACTTTATGTT ATGTAGCCTT TAGAGGCAAC AGTTCAGGAC 7380

TGGAGAGATG GCTCAAGGGT TAAGAGCACT GGCTGTTTTT TCAGAGGACC CATGTTTGAC 7440

TCACAGCACA CACATGGTGG CTCACAGCCA TCATGACTCC TGTTCCAAAG GATCTGATGT 7500

CTTCTTCTGA CCTCTGCAGA CACCAGGCAT GCATACATGC AGGCAAAATA CCCATCAATA 7560

TAAAAATAAA TAACTGGGAA ATATGCAAAT TCTTTAATAT GCAAATTCTT CTCTCCCCAA 7620

CTGCCATTTC CCATGCTCCA CCCTCATCCC TTCCCTCCTC TCTTACTTCT TTTGTTTGGA 7680

ATTCTTTAGA TAGCATCATC AAGGAGGCTC TGAGGCTTTC CAGTGCATCC TTGAATATCC 7740

GGACTGCTAA GGAGGATTTC ACTCTGCACC TTGAGGATGG CTCCTATAAC ATCCGAAAAG 7800

ACGACATCAT CGCTCTTTAT CCACAGTTAA TGCATTTGGA TCCTGCAATC TACCCAGACC 7860

CTCTGGTAAG TTTTTCTGCT CATCAAAGTT ATGTATCGAG GTGACAGTCA CCCAGGAATG 7920

TATTTGTAAT TACAGCTTTG ATTTGATCAT TAAAGTGAAG CCATAGGGAT TGTCCCTCTT 7980

TATTGCGGCA AATATTCATG TTTTGGAAAC TTTGGGTAGA GGCAAGAGTT TTGAACTTTT 8040

ACACCTAATA TTCATTTCAT AGTTTCTGCT AGACTATGTT TTCAGTCATA ACAAAACTAC 8100

CACCTTTTTT CCCCCTCACA AAGTACCCTC TCCCAAATTT ACACTAATGG AGGGTAATGC 8160

ATTTGACTTG ATCCTTAGAG TAGTTGTTTA GAGCCATTTT GCTTCTTTTG TCTAACTGAA 8220

GAATTAGTCT ACAGGTAGAA CAGGAGGTCC CTAGAGCTTC TTGGTCCACC AGCTCTTCAT 8280

AAGCTCTTTC CAGTATCACC TGGTTCAGTG CTTGGTGTTT GCTAACTTGT AGAGGATGGA 8340

TTTATTAGTA GAAAATTACT CTTTGGATCC TCCAGGTCAA GAAGGCAACA ACTTTCTATC 8400

ATAATAGCTC ATTGGCTTCT TGTCTCTTTG TTGCAGACTT TTAAATATGA TCGATACCTG 8460

GATGAGAACA AGAAGGCAAA GACCTCCTTC TATAGCAATG GAAACAAACT AAAGTATTTC 8520

TATATGCCAT TTGGATCCGG AGCTACAATA TGCCCTGGGA GACTATTTGC TGTCCAAGAA 8580

ATCAAGCAAT TTTTGATTCT GATGCTTTCA TACTTTGAAC TGGAGCTTGT GGAGAGTCAT 8640

GTCAAGTGTC CTCCTCTAGA CCAGTCCAGG GCAGGCTTGG GGATTTTGCC ACCATTAAAT 8700

GATATTGAGT TTAAATATAA ACTGAAACAT CTGTGACATG TGGTTGGAAG AAGAGGACAC 8760

TGGATGATGT TGCTGGACTG CAGCGAGTCT CACTAAACAA GCCCTTGGGA CAAATGCTCT 8820

CCTTTGCTTC CCAGCAACTG ACTGTGCCTA GGAAAAGAAC TGGTACCCCC GGCACCACTC 8880

TCTGTTCTCA CTGCCTGAGT TCCTGGGTGT TCAGATAGCT GAGGTCAGAG TTTCACCACT 8940

CTTAGAAGCA ATGTCTTTTG TTTTTATTTT CAAAATGAAG ATACTCCAAT TGGCAGATTT 9000

TTTTTCCTAA GGAAATTGCT TCATACTTTT ATGAAAACTG ATTAATTATG AAAAGGCTTC 9060

AAATTCACGT TTTAGTGAAA CTGTTATTTT TTTCACTAGT GAAGTTCTTC ATGTGTGAAC 9120

ATATACTATA AAAACATTTT AAGGGATCAT ATCATGCTTT GCATAAAGGG AAAGGAAAAT 9180

ATTATTCAAC TTTTTTTTTT GGTTTTTCTA GACAGGGTTT CTCTGTGTAG CTTTGGAGCC 9240

TATCCTGGCA CTCACTCTGT AGAGCAGGCT TGGTCTTGAA CTCACAGAGA TCTGCCTGCC 9300

TTTGCCTTCC GAGTGCTGGG ATTAAAGTCG TGCGTCACCA ATGCCTGGCT ATTTAACTTT 9360

TTCGATGTCT AGTGGTGAGA GCTTTGAAAA TGATGCTACT GTGTTGGGAA TACTATGGGA 9420

AATTTTGATG CTTCGCTGTT ACATTTAAAT TTATTGCTGC TGGAAATTGT CACCCCAGTT 9480

TTCAATTGCC CCTCTCTCTC CCTTTTAATA TTCACACTGA TGAGCAGAGT TTTTTAGAGA 9540

TTAAAAAGAC CTCCCCAGAG CCCTGTCTCT GATGTTTTTA AGCCTTTAAT CTCAGTACTC 9600

AGGAGGCAGA GGCAGGCAGA GCTCTGTGAG TTCGAGGCCA GCCTGATCTA CAGATCGAGT 9660

TCCAGGCAAG CCGGGGCTAC AGAATGAGAC CTTGTCACTA AAAGAAATAA ATAAGGTCAA 9720

TTTTATGTCA CAACTGATTA TGAATCATTG TAAAGGATAA ATTGAAAAAA AAGAACTCCA 9780

CGGGAATGAC CATTTAAATG GTCTATTTTA GCTAAAATTA ACTATGAATT ATGTGGAGTT 9840

CATTAAGTGT ATGTTGACGT TATATGTTCC TTTAAAATGT CTTATGTTTT ATCTCTGAAT 9900

```
GTCTTGTAGA TGGAGAGCAA TAATAGTGTT TAAATACTGA GTCAATAAGG TTTTATCTAT   9960
GTACTTTAAG AGCATTATTA GCTGTGTCAT TTTTACTGAT ATATCTAATA TATTTATATG  10020
TAAATTATAT TTATCTTTTA TCTTATACTA CAAATATAAG TAAATATTTT AAAACCAGTA  10080
ACTTTAAAAT TACCTACCTT TCAGAAATGA AAATAAGAAC ATTTGTGCTT TAACCTTTGA  10140
AATAGAATGT TTATTCATCC ACTGATAAGT TAAAATAATT TTATCTGATT TGTTTCAAGA  10200
AACTCAAAAA TATTCAAAGT AATCATGCAC TCAAAGGTCT TCGTAAGGTT ACAGAAAATT  10260
CAATAAAATC TTTTTTGTGT AGGGACTGAG TCAGGGTCTA GAAGATGCTT GGCAGGTACT  10320
CCAGTAGTGA GCTGGATCCA GAAGATTCCT TAAACTTTAA AATCTTAACA CTAAGTATTA  10380
TCACAGAGTT ATTACCTAAG TAGAATATTT TTCCTTTCCT TTTCAATTGA CAGAGTCCCA  10440
CAGCAACACA GCTGGCTGTA ACTCTTCACA TAGCTTGCGC AGGCTTTGAA CTCACTGTAC  10500
TCCTGCCTTT CCTTTTCTAG GAAATTATTT TCCACATCAA GAAAATTTAA TTGTTCCGAT  10560
GAGGTATAGA GTAACAAATT TCTGTTATAT ATTCATCTGT ATTAAACTGA ATTC         10614
```

Example 2. REGULATORY ELEMENTS AND TRANSCRIPTION FACTORS

Cloning of the CYP7 gene from three different species allows the analysis of the CYP7 gene structure and organization. Alignment and analysis of the highly conserved proximal promoter region of these homologous gene suggests that many regulatory elements are conserved and are likely to play important roles in gene regulation. Mapping of these transcription factor binding sites is essential to the isolation of transcription factors involving in the regulation of liver-specific CYP7 gene transcription. These sequence elements and protein factors are potential models for designing compounds and for screening for activators or repressors of the gene, such as described in a parent U.S. Application Serial No. 08/135,488, to Chiang, J.. The following discussion relates to the regulatory elements and transcription factors of the rat gene promoter.

2.1. Alignment and Analysis of the CYP7 Genes

The proximal promoter regions of the rat, human and hamster genes were aligned. Sequence identity is about 82% between rat and hamster, 77% between hamster and human and 71% between human and rat (Fig. 4). Several liver-enriched transcription factors, HNF3, HNF4, HNF1 and C/EBP, and thyroid/steroid hormone response elements are highly conserved in these homologous genes (Fig. 5). Sequences that are further upstream of these genes have diverged considerably. In contrast to the report that the -400 proximal promoter of the human gene had no promoter activity (Molowa, et al. Biochem. 31, 2539-2544, 1992), this conservation indicates that the proximal promoter is important in transcriptional activation function and contains essential regulatory elements.

2.2. Footprint Analysis of the Rat Gene

DNase I hypersensitivity sites of the rat gene were mapped by digestion of rat liver nuclei (20 OD_{260}) with DNase I at 37°C for time periods up to 4 minutes. DNA was isolated from nuclei at each time interval and digested with SacI, fractionated on a 0.8 % agarose gel and transferred to nylon membranes. A 5'-probe of Sac I-EcoRI fragment (-3643 to -2265) was used for indirect end-labeling and was labeled with an activity of at least 1 x 10^9 CPM/µg. Four DNase I hypersensitivity sites (HSI, HSII, HSIII, HSIV) were mapped. HSI is mapped near a "CA" repeat region around nucleotide-1,500. HSII is located in the proximal promoter region. HSIII and HSIV are located in intron I and intron II, respectively (Fig. 6).

DNase I footprinting technique then was applied to map the transcription factor binding sites in the gene promoter (Heberlein,U, England, B and Tjian, R Cell, 41, 965-977, 1985). Transcription factor binding sites in the gene are protected from DNase I digestion. Two fragments were mapped: a Hind III-Xba I fragment (-346 to +36) in the proximal promoter region near the hypersensitivity site II and an upstream fragment Xba

I-Hind III(-1530 to -1205) in the hypersensitivity site I. Probes were made from plasmid DNA digested with a restriction enzyme to generate a 5'-overhang, filled in with the Klenow fragment of DNA polymerase I and ^{32}P-labeled dCTP, and then digested with a second restriction enzyme. Probes were purified from a native 5% polyacrylamide gel. Footprinting reactions included 2 μg of poly(dI-dC), 10% polyvinyl alcohol, 50 mM KCl and 20fmol of probe in a volume of 50 μl. Reactions were stopped with EDTA and SDS, then phenol extracted, ethanol precipitated and run on polyacrylamide sequencing gels.

The footprinted areas are summarized as follows:

Footprints (FP) mapped in hypersensitivity site II:

FP I (Nucleotides -81 to -37): TGT3, 7α-TRE, HNF1 /LFB1, CAAT, Box elements

5'-TGTTTGCTTTGGTCACTCAAGTTCAAGTTATTGGATCATGGTCC-3'

FP II (Nucleotides -149 to -131): HNF4/LFA1 element

5'-CTATGGACTTAGTTCAAGG-3'

FP III (Nucleotides -171 to -154): GRE half site

5'-TGTTCTGGAGCCTCTTCT-3'

Footprint mapped in hypersensitivity site I:

FP IV (Nucleotides -1448 to -1410): NF1 elements

5'-TCACTGTGGCCTAGTGCCACATCTACCTATTTCTTTGGCTTTACTTTGT-3'

Footprint I covers a sequence from nucleotide -81 to -37 and consists of four elements: TGT3/HNF3, 7α-TRE, LFB1 /HNF1, and CAAT box (reversed). Footprint II covers sequence from -149 to -131 and contains an LFA1/HNF4 site. Footprint III covers sequences from -171 to -154 and contains a consensus glucocorticoid response element (GRE) half site. In the hypersensitivity site I, a footprint covers - 1554 to -1505 and contains a bipartite and a half-site of the NF1 /CTF element. Most of these sequences are liver-enriched transcription factor consensus motifs and are highly conserved in all three species. It is especially interesting that Footprint I contains overlapping binding sites for at least four transcription factors, HNF3α/3β, 7α-TRE, HNF1/LFB1, and C/EBP. The TRE-like sequence (TGGTCANNNNAGTTCA) located in the center of the cluster may be the binding site for Type II hormone receptors such as the T_3 receptor (T_3R), the retinoic acid receptor (RAR), the retinoid X receptor (RXR), the vitamin D_3 receptor (VD_3R), or the peroxisome proliferator activating receptor (PPAR) (Stunnenberg, HG, BioEssays, 15, 309-315, 1993). This gene fragment has been shown to be essential for major promoter activity and could confer taurocholate repression of promoter activity in rat primary hepatocyte cultures. It is likely that the element in footprint I identified in the present invention is a bile acid response element (BARE) of the CYP7 gene.

2.3. Gel Mobility Shift Analysis of the Rat Gene

The electrophoretic mobility shift assay (EMSA) is used to detect specific DNA-protein interactions in the identified footprints. Oligonucleotides corresponding to PPRE/TRE, 7αTRE, and TGT3 were synthesized and annealed to form double-stranded probes. DNA fragments corresponding to Footprints I, II, and IV were generated by PCR using primers that flank the footprint sequences. Probes are labeled with ^{32}P dCTP by the Klenow fragment of DNA polymerase I. Probes were gel purified before use. Binding reactions were done in 20 μl comprising 10 % glycerol, 10 mM HEPES, pH 7.9, 2 μg of poly(dI-dC), 1 μg of nuclear protein extracts and 20,000 CPM of probes at 30°C for 15 min, followed by electrophoresis on 4% native polyacrylamide gels (Carthew, RW, et al. Cell, 43, 439-448, 1985).

The footprint I probe shifted at least 4 bands when it was reacted with liver nuclear extract. Cold competitor specifically prevented band shifts. The footprint II probe shifted two bands whereas Footprint IV probe shifted only one band with liver nuclear extract. Since Footprint I contains several transcription factor binding elements and is the possible bile acid receptor or binding protein (BAR) binding site, double-stranded oligonucleotides were synthesized corresponding to the TGT3 and 7α-TRE elements in Footprint I.

EMSA revealed that the TGT3 element shifted two major bands, which may be due to the binding of HNF3α and HHF3β, whereas the 7α-TRE element shifted two different bands. Protein factors that bind to the 7α-TRE probe could be competed out with a 100-fold excess of its cold competitor or a rat growth hormone gene TRE element. However, TGT3 and PPAR/TRE oligonucleotides did not compete with the 7α-TRE probe. These results indicate that the 7α-TRE like element identified in the CYP7 gene promoter binds to one or two specific liver protein factors. In addition, the 7α-TRE of the human CYP7 gene (Figure 4) also shifted one band in human liver nuclear extracts.

Furthermore, EMSA was performed using liver nuclear extracts isolated from rats treated with a diet supplemented with 0.25% deoxycholate, 1% cholate, 5% cholestyramine or 1% cholesterol for two weeks. Only nuclear extracts from deoxycholate-treated rat liver abolished the gel shift of the 7α-TRE oligonucleotide. Deoxycholate or sodium cholate treatment reduced both cholesterol 7α-hydroxylase activity

and mRNA levels by 80% and 60%, respectively, whereas cholestyramine or cholesterol treatment stimulated these parameters by 330% and 180%, respectively.

These results suggested that deoxycholate may inhibit the binding or synthesis of a positive nuclear transcription factor, (i.e. factor A) to a bile acid responsive element (BARE) or inhibit the synthesis of factor A in nuclei as well as repress CYP7 gene expression. Alternatively, deoxycholate may bind to a negative regulator, BAR, which forms a complex with the positive factor A and prevents the binding of factor A with BARE. BAR and nuclear transcription factor A may compete for the same binding site, BARE. These factors are likely members of the steroid/thyroid hormone supergene family, since the recognition sequence is similar to the cognate response element. Interactions between this transcription BAR with adjacent liver-enriched transcription factors (HNF3α, HNF3β, HNF1, C/EBP) can affect the expression levels of the CYP7 gene.

2.3(a) Effect of Bile Acids on EMSA: Further Results

A gel shift experiment was performed to further confirm that the 7αTRE and the DR_0 elements are bile acid responsive elements. Liver nuclear extracts isolated from rats treated with dietary supplements specified above were used. Deoxycholic acid and sodium cholate treatment significantly suppressed both cholesterol 7α-hydroxylase activity and mRNA levels by about 80% and 60%, respectively. On the other hand, 5% cholestyramine or 1% cholesterol stimulated activity and mRNA level by 330% and 180 respectively.

The rat 7αTRE element shifted one band in human nuclear extracts, while the human 7αTRE shifted one band in all rat nuclear extracts that were treated with cholestyramine, sodium cholate and cholesterol. In deoxycholate-treated rat liver nuclear extracts, however, human 7αTRE did not shift any protein band. All other nuclear extracts showed similar band patterns (no shift) as that of the control (non-treated rat) extracts.

From the gel generated by this experiment, it was observed that 7αTRE shifted two bands whereas rat DR_0 shifted one. Thus, rat DR_0 element appeared to bind the transcription factor more specifically than did 7αTRE. Accordingly, the rat DR_0 element was selected for use as a probe to demonstrate the presence of transcription factor on a Southwestern blot, discussed below.

2.3(b) Characterization of a DNA-Binding Protein

A Southwestern blot, which illuminates DNA-protein interactions, was performed to reveal nuclear protein factor(s) that bind to the rat DR_0 element, which appears to bind a transcription factor more specifically than 7α-TRE. This rat probe predominantly bound to a polypeptide of about 57,000 + 7000 Daltons which showed a similar band width in all rat liver nuclear extracts tested, including extracts from non-treated rats and rats treated with cholestyramine, sodium cholate, cholesterol and deoxycholate.

The rat DR_0 revealed a second band shift of 116,000 daltons in all of these extracts as well. This second shift is believed to constitute a dimer of two 57 KDa peptides. The 57 KDa polypeptide was also present in nuclear extracts of rat spleen, rat kidney and human liver, although the band was less pronounced in the human liver extracts.

Methods of substantially isolating transcription factors according to the invention, for example, can employ DNA fragments according to the invention in conjunction with methodology taught by Singh et al., Cell 52: 415 (1988) and Kadonaga et al., PNAS USA 83: 5889 (1986). Each of these publications is incorporated by reference herein in their entirety. Yet another approach to identify and clone genes for proteins that interact with DNA-binding protein employs yeast two-hybrid system to study protein-protein interaction (Fields and Song, 1989; Chien et al. PNAS 88:1958 (1991)).

2.4 Recognition site affinity chromatography

One approach to isolating a transcription factor provides the advantage of isolating a protein complex that includes both a DNA-binding protein and other associated protein factors that interact with the a DNA-binding protein. The success of purification of transcription factor is dependent, generally, on parameters including the quality of nuclear extracts, the amount of transcription factors present in the extracts, and the binding affinity of the DNA-affinity column. Also, the binding site sequence selected for use in the column is optimized by EMSA, DNase I footprinting, mutational analysis and by sequence comparison of homologous binding site.

A BARE consensus sequence, such as that identified by the present invention can be utilized advantageously in an affinity column. The rat or human 7αTRE or DR_0, which recognized a single binding

protein in EMSA of either human or rat nuclear extracts, is suitable for DNA affinity column chromatography and identifies a 57 KDa bile acid responsive protein.

Affinity chromatography is performed according to the following protocol. First, cell-free nuclear extracts are obtained from either HepG2 cells or human liver tissues. Fresh human liver tissue is advantageous for isolating nuclear extracts, however, it sometimes difficult to obtain. Nuclear proteins are extracted with high salt and crude extracts are precipitated with ammonium sulfate, dialyzed and then subjected to gel filtration column (i.e., Sephacryl S-300, Pharmacia) or a heparin-agarose affinity column (Sigma Chemical). Column fractions are assayed for transcription factors by EMSA and pooled fractions are applied to a sequence-specific affinity column.

A DNA affinity column is prepared which employs double-stranded BARE consensus oligonucleotides according to the invention, which are provided with a 5' overhang of nucleotides "gatc". The oligonucleotides are concatemerized by phosphorylation with T4 polynucleotide kinase and ligated by T4 DNA ligase (Jackson et al., GENE TRANSCRIPTION: A PRACTICAL APPROACH, ed. Hames and Higgins, I.R.L. Press 189-242 (1993). The disclosure of the relevant section of this book concerning affinity chromatography methodology is expressly incorporated herein by reference in its entirety. The ligated, concatemerized DNA is covalently attached to a CNBr-activated Sepharose Cl2B gel (Pharmacia) (Kadonaga, 1986 supra).

A transcription factor preparation isolated by the column is subjected to SDS-polyacrylamide gel electrophoresis. Thereafter, the gel is stained with silver staining to demonstrate the preparation's purity, and the DNA-binding properties of the purified transcription factor measured using EMSA and DNase I footprinting.

Once a transcription factor is purified, it can be used to raise antibodies, which in turn are used as a screening probe to isolate cDNA clone encoding the transcription factor. For example, the purified 57 KDa BARP is used to raise antibodies against itself, which are used as a screening probe to isolate its cDNA.

2.4(a) Screening using recognition-site sequences.

An alternate method of isolating a BARP includes directly cloning cDNAs encoding a BARP from human liver cDNA expression libraries (Promega, Clontech), which are screened for a fusion protein recognizing specific nucleotide sequences. This technique is perhaps simpler than affinity chromatography, but it yields cDNA(s) that encode a DNA-binding protein, not protein itself.

Binding site probes of a BARE consensus sequence according to the invention are prepared by 5'-end labelling a double-stranded oligonucleotide with γ-^{32}P ATP using T_4 polynucleotide kinase. T4 DNA ligase is then used to concatamerize oligonucleotides. Human liver λgt11 cDNA expression libraries will be screened following routine procedure described by Sambrook et al., Mol. Cell Biol. 9:946 (1989).

Fusion proteins are induced by overlaying the plates with IPTG-treated nitrocellulose filters and incubating for 6 hours at 37 • C. Filters are soaked in 6 M guanidinium chloride in binding buffer and washed in the same buffer but gradually reducing the concentration of denaturant to 0.188 M, and finally in buffer without denaturant. Filters are placed in binding buffer, and blocked in non-fat mild solution and incubate with binding site probe at 4 • C overnight. Filters are washed and autoradiographed at -70 • C.

Positive plaques are picked, replated and screened until plaque-purified. cDNA is sequenced by dideoxy chain termination method using Sequenase Kit (USB Co.) and analyzed with DNA analysis software. Amino acid coding sequences are analyzed for sequence motifs and compared against GenBank database for characteristics of DNA-binding proteins, such as possessed by a zinc finger, leucine zipper or member of a nuclear receptor gene family.

2.5 Characterizing transcription factors

To overexpress a BARP for footprinting and transient transection assays, its cDNA is isolated according to the protocol of 2.3(b) and subcloned into a pMT eukaryotic expression vector (Kaufman et al., 1989). For gel shift assay, cDNA will be subcloned into pGEM4 (Promega). Plasmid is linearized and in vitro transcribed by SP6 RNA polymerase. The resulting RNA is translated in a rabbit reticulocytes lysate system in the presence of ^{35}S-methionine.

EMSA is performed as described herein. In vitro synthesized protein is incubated with ^{32}P-labeled probe and electrophoresed in low ionic strength polyacrylamide gel. Two filters are placed against the dried gel, the first of which blocks the ^{35}S radiation.

CYP7 promoter/luciferase constructs and pMT plasmid carrying a BARP cDNA are transiently cotransfected into HepG2 cells by calcium phosphate coprecipitation method as described previously. pRSV-βgal

plasmid is used as an internal standard for normalization of transection efficiency. A test agent or endogenous factor is added in culture media and incubated for a period of time. Cells are lysed, then luciferase activity is measured, as described previously.

Electrophoretic Mobility Shift Assay of DNA-protein Interactions

Sequences of double-stranded probes	# of bands shifted
1). FP I probe (-100 to -29):	four to five

```
5'-
CTAGTAGGAGGACAAATAGTGTTTGCTTTGGTCACTCAAGTTCAAGTTATTGGATCATGGTCC-3'
GATCATCCTCCTGTTTATCACAAACGAAACCAGTGAGTTCAAGTTCAATAACCTAGTACCAGG-5'
3'-
```

2). FP II probe (-161 TO -127): two

```
5'-CCTCTTCTGAGACTATGGACTTAGTTCAAGGCCGG-3'
3'-GGAGAAGACTCTGATACCTGAATCAAGTTCCGGCC-5'
```

3). FP IV probe (-1454/-1394): one

```
5'-TCACTGTGGCCTAGTGCCACATCTACCTATTTCTTTGGCTTTACTTTGTGCTAGGTGACC-3'
3'-AGTGACACCGGATCACGGTGTAGATGGATAAAGAAACCGAAATGAAACACGATCCACTGG-5'
```

4). PPRE/TRE element probe (nt -101/-82): two

```
5'-GAAGATCTAGTAGGAGGACAAATAG        3'
3'          CATCCTCCTGTTTATCAC      5'
```

5). 7α-TRE element probe (nt -73/-56 in FP I) : two

```
5'-GATCCTTGGTCACTCAAGTTC        3'
3'     GAACCAGTGAGTTCAAGTTCCTAG  5'
```

6). TGT3 element probe (nt -86/-71 in FP I) : two

```
5'-GATCCAATAGTGTTTGCTTTGGT          3'
3'          TCACAAACGAAACCATCCTAG   5'
```

2.6 Promoter/Reporter Gene Constructs

To determine the promoter sequences responsible for regulation of cholesterol 7α-hydroxylase, deletions of the rat CYP7 promoter were ligated upstream of the luciferase reporter gene (Luc). The promoter fragments were generated by the polymerase chain reaction using the primers listed with a rat CYP7 genomic clone as the template. The fragments were blunted by filling in with the Klenow fragment of DNA polymerase and then digested with Xho I. The fragments were then ligated into the pGL2-basic vector (Promega) which had been digested with SmaI and Xho I, and transformed into E. coli HB101 cells. The resulting plasmids (pLUC-224, pLUC-160, pLUC-101, and pLUC-3600) are used to transfect primary hepatocytes or hepatoma cells for the study of luciferase gene expression under the control of the CYP7 promoter. The results show that pLUC-224 had two-fold higher luciferase activity than pLUC-160 and pLUC-3600 when transfected into rat primary hepatocytes. pLUC-3600 had transcription activity similar to that of pLUC-160. In addition, 50μM taurocholate inhibited the expression of luciferase activity in these hepatocytes, indicating that these CYP7 gene promoter fragments do contain a BARE, which confers bile acid regulation.

To determine if the sequence from -101 to -29 of the CYP7 gene promoter can function as an enhancer element, the region was cloned into the pGL2-Promoter vector (Promega). The vector is similar to pGL2-basic, with the addition of the SV40 early promoter between the multiple cloning site and the luc gene. The rat sequence was amplified by the polymerase chain reaction to flank the sequence with a BamHI site and a BglII. The fragment was ligated in both orientations to the pGL2-Promoter, which had been cleaved with BglII. The resulting plasmids are named pLUC-101/-29 and pLUC-29/-101.

Chloramphenicol acetyltransferase (CAT) reporter gene constructs were made by using the polymerase chain reaction and primers to amplify the region -415 to +36 of the rat CYP7 gene and to incorporate an XbaI at nucleotide + 36. The blunt ended, Xba I digested fragment was ligated into a promoter-less pCAT basic vector (Promega) which had been digested with Sal I, blunt-ended and digested with Xba I to yield -415CAT. A longer construct named -3643CAT was made by digesting - 415CAT with Hind III and inserting a 3.2 kb Sac I-Hind III genomic fragment. The 3.6 kb insert was removed from -3643CAT and ligated into a pGL2-basic vector (Promega). This plasmid was used to generate nested deletions with Exo III and S1 nuclease.

Promoter/Reporter Gene Constructs

PCR primers used for PCR of fragments

```
                          +30
L1: 5'-AGATGGCTCGAGACTCTTTGCCTAGCAAA-3'
                      XhoI

          -224
L3: 5'-CAGCACATGAGGGACAG-3'
          -160

L4: 5'-CTCTTCTGAGACTATGGAC-3'
                -101

L8: 5'-GAAGATCTAGTAGGAGGACAAATAG-3'
             BglII
```

Sequences of promoter fragments inserted in pGL2-basic vector

pLUC-224:

```
5'-CAGCACATGAGGGACAGACCTTCAGCTTATCGAGTATTGCAGCTCTCTGTTT
   GTTCTGGAGCCTCTTCTGAGACTATGGACTTAGTTCAAGGCCGGGTAATGCTATT
   TTTTTCTTCTTTTTTCTAGTAGGAGGAGGACAAATAGTGTTTGCTTTGGT
   CACTCAAGTTCAAGTTATTGGATCATGGTCCTGTGCACATATAAAGTCTAGTCAGA
   CCCACTGTTTCGGGACAGCCTTGCTTTGCTAGGCAGGCAAAGAGTCTCGAG-3'
                                                  XhoI
```

pLUC-160:

```
5'-CTCTTCTGAGACTATGGACTTAGTTCAAGGCCGGGTAATGCTATTTTTTTCT
   TCTTTTTTCTAGTAGGAGGACAAATAGTGTTTGCTTTGGTCACTCAAGTTCA
   AGTTATTGGATCATGGTCCTGTGCACATATAAAGTCTAGTCAGACCCACT
   GTTTCGGGACAGCCTTGCTTTGCTAGGCAGGCAAAGAGTCTCGAG-3'
                                              XhoI
```

pLUC-101:

5'-GAAGATCTAGTAGGAGGACAAATAGTGTTTGCTTTGGTCACTCAAGTTCA
AGTTATTGGATCATGGTCCTGTGCACATATAAAGTCTAGTCAGACCCACT
GTTTCGGGACAGCCTTGCTTTGCTAGGCAGGCAAAGAGTCTCGAG-3'
XhoI

pLUC-3600:

3.6 kb 5' flanking sequence to +36

Sequences of promoter fragments inserted in pGL2-promoter vector:

---

pLuc-101-/-29:

-101
GAAGATCTAGTAGGAGGACAAATAGTGTTTGATTTGGTCACTCAAGTTC
-29
AAGTTATTGGATCATGGTCCTGTGCACATCCTAGGGC-3'

pLuc-29/-101:

Reversed direction of the above sequence

Promoter/CAT gene constructs:

-415CAT:

sequence from -415 to +36

-3643CAT:

3.6 kb 5'-upstream sequence to +36

Example 3. HepG2 CELLS TRANSFECTED WITH PROMOTER/REPORTER GENE CONSTRUCTS

3.1 HepG2 cell cultures

HepG2 cells were obtained from ATCC (Bethesda, MD) and grown in Dulbecco's Modified Eagles Medium/F12 (50:50) supplemented with 10% heat inactivated fetal bovine serum, 1 mM Minimum Essential Medium (MEM) sodium pyruvate, 1 x MEM non-essential amino acids, 25 mM Hepes, 100 U/ml penicillin G and 100 mg/ml streptomycin in a humidified incubator with 5% CO_2 in air at 37 • C. Forty-eight hours prior to the isolation of RNA, the media were replaced with fresh media containing bile salts but without fetal calf serum. The monolayers were grown to either subconfluent (50 to 70% confluent) or confluent. Viability of the cells was checked by Trypan Blue exclusion test. About 40 million cells were lysed by the addition of 4 M guanidinium thiocyanate, 0.5% N-lauroylsarcosine, 25 mM sodium citrate, pH 7.0, and 0.1 M 2-mercaptoethanol, phenol extracted, and ethanol precipitated. Poly (A +) RNA was isolated using PolyAT tract mRNA isolation system III according to the manufacturer's instructions (Promega, Madison, WI). A PstI fragment of human cholesterol 7α-hydroxylase cDNA was labeled with ^{32}P and used as a hybridization probe, according to the method of Karam et al., Biochem. Biophys. Res. Commun. 185: 588 (1992). Human actin cDNA was used to hybridize the same membrane and served as an internal standard for the normalization of RNA level which were quantitated by scanning each lane with a laser scanner.

For transient transfection assay, cells were split and plated for at a density of 10^5 cells/60 mm Petri dish and grown to subconfluence (about 30% confluence) or to confluence.

3.2 Clarifications concerning the rat CYP7 promoter/reporter gene constructs utilized

Constructs pLUC-3600, pLUC-224, and pLUC-160 were constructed according to the description in section 2.4 above, with the following minor corrections to their nomenclature noted for the sake of exactness. First, as shown in the sequences listed above, the promoter sequences of all three constructs share the common endpoint, nucleotide +32 of the CYP7 DNA sequence, as opposed to nucleotide +36. The latter 4 nucleotides upstream of +32 include non-CYP7 bases that are a part of the exogenously-added Xho splice site. Second, as stated above, construct pLUC-3600 comprises a fragment encompassing the entire 3643 kilobase promoter region up until +32 of CYP7 gene. Accordingly this construct known figuratively as pLUC-3600 denotes a construct that contains a fragment between -3643 and +32 of CYP7.

Accordingly, three chimeric gene constructs, pLUC-3600, pLUC-224, and pLUC-160, represent deletion mutants generated by PCR using primers or by restriction digestion as described herein above. These three constructs were used for transient transfection assays in HepG2 cells.

3.3 Characterization of cholesterol 7α-hydroxylase mRNA in HepG2 cells.

HepG2 liver cells express cholesterol 7α-hydroxylase normally, which makes these cells good candidates for the study of CYP7 regulation. To assess the suitability of these cell lines suitable for use in a transfection assay of the CYP7/reporter chimeric gene constructs, it was necessary to prove first that cholesterol 7α-hydroxylase activity could be regulated in these cells.

To characterize HepG2 cells, expression of cholesterol 7α-hydroxylase mRNA was measured in HepG2 control cells and cells treated with bile acids. Northern blot hybridization of poly (A+) RNAs isolated from confluent cultures of HepG2 cells, that were treated with media containing 100 μM of tauro-(T) or glyco-(G) conjugate of cholate (CA), deoxycholate (DCA), chenodeoxycholate (CDCA) or ursodeoxycholate (UDCA) and incubated. Cholesterol 7α-hydroxylase cDNA hybridized to two mRNA species of 3 kb and 1.8 kb, in agreement with Hassan et al., Biochem. Pharmacal. 44: 1475 (1992). Both of these RNA species apparently are 7α-hydroxylase mRNA because the two bands changed responsively in parallel.

In subconfluent cultures, only TCDCA could repress mRNA level. In contrast, tauroursodeoxycholate (TUDCA) significantly increased mRNA level in subconfluent HepG2 cells. Glyco-conjugates of bile acids had similar effects as the tauro-conjugates. At this concentration, bile acid did not reduce viability of HepG2 cells.

Figure 7 summarizes the effects of bile acid conjugates on 7α-hydroxylase mRNA level in HepG2 cells. When 100 μM taurocholate (TCA) was added, mRNA level was not changed significantly, while TDCA and TCDCA reduced mRNA level by 50 to 80% in confluent cultures. mRNA levels are expressed as % of mRNA level in cells without treatment of bile acids. Values are averages of three experiments. Thus, cholesterol 7α-hydroxylase mRNA level in HepG2 cells is regulated by bile acids. The inhibitory effect of bile acids follows the hydrophobicity indexes of bile acids, TCA < TDCA < TCDCA, as described by

Heuman et al., Lipid Res. 30: 1160 (1989). The results are also consistent with those observed in primary cultures of rat hepatocytes, as described by Hylemon et al., J. Biol. Chem. 267: 16866 (1992).

3.4 Transient Transfection of HepG2 cells with rat CYP7 promoter/reporter constructs

CYP7 promoter/reporter constructs were transiently transfected into HepG2 cells using the calcium phosphate-DNA coprecipitation method, with 0.5 ml of coprecipitate containing 5 μg of test plasmid (pLUC-3600, pLUC-224, and pLUC-160) and 1 μg of β-galactosidase expression plasmid, pCMVβ (Clontech), as an internal standard for transfection efficiency. After 4 hours, cells were shocked with 15% glycerol in TBS for 90 seconds, washed three times with TBS and further incubated for 42 hours in serum free medium containing 200 μM tauro-conjugates of bile acids. Cells were washed twice with phosphate-buffered saline, lysed and harvested with 400 μl of reporter lysis buffer (Promega) according to manufacturer's instruction.

Luciferase activity was assayed by mixing 20 μl of cell extracts to 100 μl of luciferase assay reagent (Promega) at room temperature and measuring light emission during the initial 10 seconds of the reaction. A luminometer (Lumat LB9501, Berthold) was used for this purpose. Luciferase activity was corrected for transfection efficiency.

3.5 Results: Transcriptional activity of CYP7 promoter/reporter constructs in HepG2 cultures

The promoter/reporter chimeric gene constructs according to the invention were transiently transfected into HepG2 cells to demonstrate the effect of bile acids on transcriptional activity. The untreated cells shown in Figure 14 reveal that promoter activity of pLUC-224 was much higher than pLUC-3600, and pLUC-160. Enhancer activity therefore is believed to be located between nucleotides -224 and -160. In addition, a repressor is believed to be located upstream of nucleotide -224, between nucleotides -224 and -3643.

The hormone response elements are likely located upstream of nucleotide -224, according to the following experiment. Addition of 1 μM thyroid hormone, T_4 and 0.1 μM dexamethasone increased transcriptional activity of pLUC-3600 by 2.5-fold in confluent cultures. However, this same amount of thryoid hormone and dexamethasone decreased the activity of pLUC-160 by 40% in subconfluent cultures, and had little effect on pLUC-224. Luciferase activity in each transfection experiment was expressed as % of activity in cells transfected with pGL2-control plasmid.

That bile acid response elements are located in the proximal promoter region, nucleotides -160 to +32, and also in region upstream of nucleotide -224 was revealed by the following experiment. Addition of 200 μM TCA, TDCA or TCDCA did not affect transcriptional activity of the promoter/reporter constructs transfected into subconfluent HepG2 cultures, as shown in Figure 15B. Luciferase activity in transfected cells was expressed as % of activity in transfected cells without treatment with bile acids. However, in the confluent cells, TDCA and TCDCA repressed transcriptional activity of p-pLUC-3600 by more than 70% and repressed activity of pLUC-224, or pLUC-160 by up to 45% (Figure 9A). TCA, however, did not affect transcriptional activities of these gene constructs in HepG2 cultures.

It will be apparent to those skilled in the art that various modifications and variations can be made to the compositions of matter and processes of this invention. In particular, various kinds of screening assays are encompassed that employ human CYP7 regulatory elements or its analogs. Thus, it is intended that the present invention cover the modifications and variations provided they fall within the scope of the appended claims and their equivalents.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: Northeastern Ohio Universities
(B) STREET: -
(C) CITY: Rootstown
(D) STATE: Ohio
(E) COUNTRY: USA
(F) POSTAL CODE (ZIP): 44272
(G) TELEPHONE: -
(H) TELEFAX: -
(I) TELEX: -

(ii) TITLE OF INVENTION: Cholesterol 7α-Hydroxylase Gene Regulatory Elements and Transcription Factors

(iii) NUMBER OF SEQUENCES: 27

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 504 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: human

(ix) FEATURE:
(A) NAME/KEY: Protein
(B) LOCATION: 1..481
(D) OTHER INFORMATION: /note= "Cholesterol 7α-Hydroxylase"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```
Met Met Thr Thr Ser Leu Ile Trp Gly Ile Ala Ile Ala Ala Cys Cys
1               5                  10                  15

Cys Leu Trp Leu Ile Leu Gly Ile Arg Arg Arg Gln Thr Gly Glu Pro
            20                  25                  30

Pro Leu Glu Asn Gly Leu Ile Pro Tyr Leu Gly Cys Ala Leu Gln Phe
        35                  40                  45

Gly Ala Asn Pro Leu Glu Phe Leu Arg Ala Asn Gln Arg Lys His Gly
    50                  55                  60
```

```
His Val Phe Thr Cys Lys Leu Met Gly Lys Tyr Val His Phe Ile Thr
65                  70                  75                  80

Asn Pro Leu Ser Tyr His Lys Val Leu Cys His Gly Lys Tyr Phe Asp
                85                  90                  95

Trp Lys Lys Phe His Phe Ala Thr Ser Ala Lys Ala Phe Gly His Arg
            100                 105                 110

Ser Ile Asp Pro Met Asp Gly Asn Thr Thr Glu Asn Ile Asn Asp Thr
        115                 120                 125

Phe Ile Lys Thr Leu Gln Gly His Ala Leu Asn Ser Leu Thr Glu Ser
    130                 135                 140

Met Met Glu Asn Leu Gln Arg Ile Met Arg Pro Pro Val Ser Ser Asn
145                 150                 155                 160

Ser Lys Thr Ala Ala Trp Val Thr Glu Gly Met Tyr Ser Phe Cys Tyr
                165                 170                 175

Arg Val Met Phe Glu Ala Gly Tyr Leu Thr Ile Phe Gly Arg Asp Leu
            180                 185                 190

Thr Arg Arg Asp Thr Gln Lys Ala His Ile Leu Asn Asn Leu Asp Asn
        195                 200                 205

Phe Lys Gln Phe Asp Lys Val Phe Pro Ala Leu Val Ala Gly Leu Pro
    210                 215                 220

Ile His Met Phe Arg Thr Ala His Asn Ala Arg Glu Lys Leu Ala Glu
225                 230                 235                 240

Ser Leu Arg His Glu Asn Leu Gln Lys Arg Glu Ser Ile Ser Glu Leu
                245                 250                 255

Ile Ser Leu Arg Met Phe Leu Asn Asp Thr Leu Ser Thr Phe Asp Asp
            260                 265                 270

Leu Glu Lys Ala Lys Thr His Leu Val Val Leu Trp Ala Ser Gln Ala
        275                 280                 285

Asn Thr Ile Pro Ala Thr Phe Trp Ser Leu Phe Gln Met Ile Arg Asn
    290                 295                 300

Pro Glu Ala Met Lys Ala Ala Thr Glu Glu Val Lys Arg Thr Leu Glu
305                 310                 315                 320

Asn Ala Gly Gln Lys Val Ser Leu Glu Gly Asn Pro Ile Cys Leu Ser
                325                 330                 335

Gln Ala Glu Leu Asn Asp Leu Pro Val Leu Asp Ser Ile Ile Lys Glu
            340                 345                 350

Ser Leu Arg Leu Ser Ser Ala Ser Leu Asn Ile Arg Thr Ala Lys Glu
        355                 360                 365

Asp Phe Thr Leu His Leu Glu Asp Gly Ser Tyr Asn Ile Arg Lys Asp
    370                 375                 380

Asp Ile Ile Ala Leu Tyr Pro Gln Leu Met His Leu Asp Pro Glu Ile
385                 390                 395                 400
```

```
Tyr Pro Asp Pro Leu Thr Phe Lys Tyr Asp Arg Tyr Leu Asp Glu Asn
                405                 410                 415

Gly Lys Thr Lys Thr Thr Phe Tyr Cys Asn Gly Leu Lys Leu Lys Tyr
            420                 425                 430

Tyr Tyr Met Pro Phe Gly Ser Gly Ala Thr Ile Cys Pro Gly Arg Leu
        435                 440                 445

Phe Ala Ile His Glu Ile Lys Gln Phe Leu Ile Leu Met Leu Ser Tyr
    450                 455                 460

Phe Glu Leu Glu Leu Ile Glu Gly Gln Ala Lys Cys Pro Pro Leu Asp
465                 470                 475                 480

Gln Ser Arg Ala Gly Leu Gly Ile Leu Pro Pro Leu Asn Asp Ile Glu
                485                 490                 495

Phe Lys Tyr Lys Phe Lys His Leu
            500
```

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 503 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: rat

(ix) FEATURE:
(A) NAME/KEY: Protein
(B) LOCATION: 1..481
(D) OTHER INFORMATION: /note= "Cholesterol 7α-Hydroxylase"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```
Met Met Thr Ile Ser Leu Ile Trp Gly Ile Ala Val Leu Val Ser Cys
1               5                   10                  15

Cys Ile Trp Phe Ile Val Gly Ile Arg Arg Arg Lys Ala Gly Glu Pro
            20                  25                  30

Pro Leu Glu Asn Gly Leu Ile Pro Tyr Leu Gly Cys Ala Leu Lys Phe
        35                  40                  45

Gly Ser Asn Pro Leu Glu Phe Leu Arg Ala Asn Gln Arg Lys His Gly
    50                  55                  60

His Val Phe Thr Cys Lys Leu Met Gly Lys Tyr Val His Phe Ile Thr
65                  70                  75                  80
```

```
Asn Ser Leu Ser Tyr His Lys Val Leu Cys His Gly Lys Tyr Phe Asp
                85                  90                  95

Trp Lys Lys Phe His Tyr Thr Thr Ser Ala Lys Ala Phe Gly His Arg
            100                 105                 110

Ser Ile Asp Pro Asn Asp Gly Asn Thr Thr Glu Asn Ile Asn Asn Thr
        115                 120                 125

Phe Thr Lys Thr Leu Gln Gly Asp Ala Leu Cys Ser Leu Ser Glu Ala
    130                 135                 140

Met Met Gln Asn Leu Gln Ser Val Met Arg Pro Pro Gly Leu Pro Lys
145                 150                 155                 160

Ser Lys Ser Asn Ala Trp Val Thr Glu Gly Met Tyr Ala Phe Cys Tyr
                165                 170                 175

Arg Val Met Phe Glu Ala Gly Tyr Leu Thr Leu Phe Gly Arg Asp Ile
            180                 185                 190

Ser Lys Thr Asp Thr Gln Lys Ala Leu Ile Leu Asn Asn Leu Asp Asn
        195                 200                 205

Phe Lys Gln Phe Asp Gln Val Phe Pro Ala Leu Val Ala Gly Leu Pro
    210                 215                 220

Ile His Leu Phe Lys Thr Ala His Lys Ala Arg Glu Lys Leu Ala Glu
225                 230                 235                 240

Gly Leu Lys His Lys Asn Leu Cys Val Arg Asp Gln Val Ser Glu Leu
                245                 250                 255

Ile Arg Leu Arg Met Phe Leu Asn Asp Thr Leu Ser Thr Phe Asp Asp
            260                 265                 270

Met Glu Lys Ala Lys Thr His Leu Ala Ile Leu Trp Ala Ser Gln Ala
        275                 280                 285

Asn Thr Ile Pro Ala Thr Phe Trp Ser Leu Phe Gln Met Ile Arg Ser
    290                 295                 300

Pro Glu Ala Met Lys Ala Ala Ser Glu Glu Val Ser Gly Ala Leu Gln
305                 310                 315                 320

Ser Ala Gly Gln Glu Leu Ser Ser Gly Gly Ser Ala Ile Tyr Leu Asp
                325                 330                 335

Gln Val Gln Leu Asn Asp Leu Pro Val Leu Asp Ser Ile Ile Lys Glu
            340                 345                 350

Ala Leu Arg Leu Ser Ser Ala Ser Leu Asn Ile Arg Thr Ala Lys Glu
        355                 360                 365

Asp Phe Thr Leu His Leu Glu Asp Gly Ser Tyr Asn Ile Arg Lys Asp
    370                 375                 380

Asp Met Ile Ala Leu Tyr Pro Gln Leu Met His Leu Asp Pro Glu Ile
385                 390                 395                 400

Tyr Pro Asp Pro Leu Thr Phe Lys Tyr Asp Arg Tyr Leu Asp Glu Ser
                405                 410                 415
```

```
Gly Lys Ala Lys Thr Thr Phe Tyr Ser Asn Gly Asn Lys Leu Lys Cys
                420                 425                 430

Phe Tyr Met Pro Phe Gly Ser Gly Ala Thr Ile Cys Pro Gly Arg Leu
        435                 440                 445

Phe Ala Val Gln Glu Ile Lys Gln Phe Leu Ile Leu Met Leu Ser Cys
    450                 455                 460

Phe Glu Leu Glu Phe Val Glu Ser Gln Val Lys Cys Pro Pro Leu Asp
465                 470                 475                 480

Gln Ser Arg Ala Gly Leu Gly Ile Leu Pro Pro Leu His Asp Ile Glu
                485                 490                 495

Phe Lys Tyr Lys Leu Lys His
            500
```

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 504 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: hamster

(ix) FEATURE:
(A) NAME/KEY: Protein
(B) LOCATION: 1..481
(D) OTHER INFORMATION: /note= "Cholesterol 7α-Hydroxylase"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

```
Met Met Thr Ile Ser Leu Ile Trp Gly Ile Ala Met Val Val Cys Cys
1               5                   10                  15

Cys Ile Trp Val Ile Phe Asp Arg Arg Arg Arg Lys Ala Gly Glu Pro
            20                  25                  30

Pro Leu Glu Asn Gly Leu Ile Pro Tyr Leu Gly Cys Ala Leu Lys Phe
        35                  40                  45

Gly Ser Asn Pro Leu Glu Phe Leu Arg Ala Asn Gln Arg Lys His Gly
    50                  55                  60

His Val Phe Thr Cys Lys Leu Met Gly Lys Tyr Val His Phe Ile Thr
65                  70                  75                  80

Asn Ser Leu Ser Tyr His Lys Val Leu Cys His Gly Lys Tyr Phe Asp
                85                  90                  95
```

```
Trp Lys Lys Phe His Tyr Thr Thr Ser Ala Lys Ala Phe Gly His Arg
                100                 105                 110

Ser Ile Asp Pro Asn Asp Gly Asn Thr Thr Glu Asn Ile Asn Asn Thr
        115                 120                 125

Phe Thr Lys Thr Leu Gln Gly Asp Ala Leu His Ser Leu Ser Glu Ala
    130                 135                 140

Met Met Gln Asn Leu Gln Phe Val Leu Arg Pro Pro Asp Leu Pro Lys
145                 150                 155                 160

Ser Lys Ser Asp Ala Trp Val Thr Glu Gly Met Tyr Ala Phe Cys Tyr
                165                 170                 175

Arg Val Met Phe Glu Ala Gly Tyr Leu Thr Leu Phe Gly Arg Asp Thr
            180                 185                 190

Ser Lys Pro Asp Thr Gln Arg Val Leu Ile Leu Asn Asn Leu Asn Ser
        195                 200                 205

Phe Lys Gln Phe Asp Gln Val Phe Pro Ala Leu Val Ala Gly Leu Pro
    210                 215                 220

Ile His Leu Phe Lys Ala Ala His Lys Ala Arg Glu Gln Leu Ala Glu
225                 230                 235                 240

Gly Leu Lys His Glu Asn Leu Ser Val Arg Asp Gln Val Ser Glu Leu
                245                 250                 255

Ile Arg Leu Arg Met Phe Leu Asn Asp Thr Leu Ser Thr Phe Asp Asp
            260                 265                 270

Met Glu Lys Ala Lys Thr His Leu Ala Ile Leu Trp Ala Ser Gln Ala
        275                 280                 285

Asn Thr Ile Pro Ala Thr Phe Trp Ser Leu Phe Gln Met Ile Arg Ser
    290                 295                 300

Pro Asp Ala Leu Arg Ala Ala Ser Glu Glu Val Asn Gly Ala Leu Gln
305                 310                 315                 320

Ser Ala Gly Gln Lys Leu Ser Ser Glu Gly Asn Ala Ile Tyr Leu Asp
                325                 330                 335

Gln Ile Gln Leu Asn Asn Leu Pro Val Leu Asp Ser Ile Ile Lys Glu
            340                 345                 350

Ala Leu Arg Leu Ser Ser Ala Ser Leu Asn Ile Arg Thr Ala Lys Glu
        355                 360                 365

Asp Phe Thr Leu His Leu Glu Asp Gly Ser Tyr Asn Ile Arg Lys Asp
    370                 375                 380

Asp Ile Ile Ala Leu Tyr Pro Gln Leu Met His Leu Asp Pro Ala Ile
385                 390                 395                 400

Tyr Pro Asp Pro Leu Thr Phe Lys Tyr Asp Arg Tyr Leu Asp Glu Asn
                405                 410                 415

Lys Lys Ala Lys Thr Ser Phe Tyr Ser Asn Gly Asn Lys Leu Lys Tyr
            420                 425                 430
```

Phe Tyr Met Pro Phe Gly Ser Gly Ala Thr Ile Cys Pro Gly Arg Leu
435 440 445

Phe Ala Val Gln Glu Ile Lys Gln Phe Leu Ile Leu Met Leu Ser Tyr
450 455 460

Phe Glu Leu Glu Leu Val Glu Ser His Val Lys Cys Pro Pro Leu Asp
465 470 475 480

Gln Ser Arg Ala Gly Leu Gly Ile Leu Pro Pro Leu Asn Asp Ile Glu
485 490 495

Phe Lys Tyr Lys Leu Lys His Leu
500

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7997 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: YES

(vi) ORIGINAL SOURCE:
(A) ORGANISM: rat

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: Clontech. RL 1022j

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 1..7997
(D) OTHER INFORMATION: /note= "Cholesterol 7α-Hydroxylase"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

GAGCTCTACC CTTGCTCTGC TATTGTACTT TTTAATACAC AGTTCAATCA AATGTGCCAC 60

CAGAATATGC ATGCTAACAG CTGTAGTGGT TGATTTTTCT TTCTACTCTT CTGTGTGTAA 120

GACCCCATGT TTTATCAATT ATTTTTTAAT GATTTCTTTC TTCATGCATA TGTGTGGTTG 180

TCAGTGTGAG TCTGTGTGTA CAGCAGGTGC ACAGGTATCC ACAGAGGCCA GAGGTTCCCT 240

GTAACTAGAA TTACAGGCAC TTGTGAACTT TCCTGTATGG GTGCTGGGAA GCAATCTGAG 300

GTCTTCTGCA AGGGATCTTA ACCACTGACT TTCTAGCCTG CTTTGCCCAT TTCTATTTAT 360

GATGACTGGA AACTGGGCTT AGGCCTTATA TTCTCTGAGG CCAAAATCAA GTTCTTCCAA 420

ACTGCAGGAT TTATGGTCTT CTATAGTATC CCACAGAAAT GGAAAAGAAA GTGACCCATT 480

AGAGCAGTAT TAGAGTCGAA ATAAACTCAA CTTGGTATGC CAGGACTTTG GACAATAATA 540

ACCCTGTCTT TTCAGGGCAT CTATCTGTAC TGCTGCAATA GAAACTCCAC AGGTCAGGGT 600

CACAGCTGTT GTGTTTTACA CAGTGTCCCC AGGATTAGTT CAGTGCCCAC CATGCAATAG 660

GTGTCATGGT GTGTGTGTGT GTGTGTGTGC GTGTGTCGTG CTTGTGTGCA TGTGTGTGAG 720

ACACACACAC AGAGAGATAC AAAGACAGAA ACAGAAAATT AATAAAATTT TACCAACTAA 780

AATAGGGAAT TAAAGAAAAG GAGGAGAAAA AGTTGGGCAT TCAACACCAT AAAGTCCCAG 840

TACTATGCTA AGAACACCCA GCTGTCCTCA CACCCGGGCA TGAAACTTCA TGCACTGTTC 900

ATCAGAAAAT CGTTTACACA CATCCCCTTG CAGTCTACTT GTAGTTTTAA CAACTTCAGA 960

GAGCACTAGC ATTTCCAGCC CCAGGTTAGA AGCTTTGGTA GATGCTGTTT GCGAGCACAG 1020

GATAGCAGCA AGAAGTGGAC TTGTTAGAAG GAAAGCCAAT GCCTATGTAA CAACGAAAAC 1080

TAAGTATGAA TCTCGAATCT CCACTCTCGT GTGTCTGTGT CTCCATATAC GTGCTTGGGT 1140

GCCTGACATG GCAAGGTGTT ACAAGTAAGG GAGGAACAAG AAAAGGACAG GGTAGTGGAC 1200

ATCAGGATGA ATGCCAGCCA GGGCGACTGG AGAGAGTCTA CGCTGCTCTG AAGGTGGGTG 1260

AAGAAGACCT CAGGAAGCTT TCTGAGGCTC CGAGAGTGCT TTTCCCTTCC CATGTTGAAA 1320

CATCCTTATT TGCAGAGAAT TCCAGGTTCA TGGGAATTTG TAAAGAGAAT ACTAAGAGGC 1380

CACCTGTGGC TTCTCCTATT TTTGTCTGCT GTCATTTATG GGACAGGGTT AGAGACCTGG 1440

CTTGCTTGGC TATGAGGCTG TTGCTTCCTC GGTTACTCTG CTGTGGTTGG ATGCATTAGG 1500

GTTAGGCCCC TCAAGAGCCA TGTGTCATTT TATAAAAGCA ATATAAATAT ACTTAAGGTG 1560

CACAAAGCAT TAGGAGGTCT GAGATAATAG ATTCTGAGAA AATCTATCCT GCTGTGTAGC 1620

AACTGATGTT TATGATTATA GTCCCAGACC ACACGATAAA GGATCTGTGG ACTCTGTTTA 1680

GGGAGGTCAA AAAACTATTG CAAATGGAGT CTATAGAGAA AACTAGACAG GACTCAATGC 1740

TCACCAATCG AGAATTAGTT GATGAGCTGG GGTAGTGACT TAGTGGATAA GAACACGGTC 1800

CTTTCAGAGG TCCTGAGTTA AATCCCCAGC AAACACATGG TGGCTCATAA CCATCTATAT 1860

TGTGATTTGA TGCCCTCTTC TGGCATGCAG GTGTACATGC AGACTCGTAT ACATAAAATA 1920

AATAAATCTT GAAAAAATGA ATACGTTGAA TAAGTGTCCC CTCGGATAAC TTTCTGCAGA 1980

ATTTTAAGCA CATGTCAATG GTAATAACAC ACACACACAC ACACACACAC ACACACACAC 2040

ACACACATAC ACACACCATA CAGATATGTA TCTAGAGACA TACACATGTA CATTTTATCT 2100

CTTTTATTTT CTTCTCCCCT CTTTGACATC AAGGAATAGA ATGCACTCAC TGTGGCCTAG 2160

TGCCACACTC TACCTATTTC TTTGGCTTTA CTTTGTGCTA GGTGACCCGA AAGGTTTAAA 2220

TATCAAAAAT GCTAATGGCT CGACATTTAC ATCCCCAATT TCTCCTTTCT CCTTACCTCA 2280

GACTCTTACA TTCAGTTGAC AATTTGACAT CGTCTCCTGG ATTTTCAAAT GTTCAGCACA 2340

CTGTACTGAT GTACTGCCTT CCAAGGCAAC CGGCACGATC CTCTCCCCAC TCCCAAGCAT 2400

CCCTCCATGA GCCAGTGTTT GCTTATCTTC TTGACTCTTG TTTTAACCCA ACTCCTCCCC 2460

TATTCACTCT GCTCTAATTC ATTCATTCTA TATTTTCGCA CATCAGGCTC ATCCTTTGCT 2520

CAGGAACTTC ACTTTTGCTT TCCGGTCTCC TGGAAATGTG TTTTCTTGGC TATTCCATCT 2580

CAAGACCATC TTTTCAGAAA AGCTTTTCCT ATCAACATAT TTAAAGCCCT CTTCATCCCC 2640

CAGTAGCTCT GGACACCTCA TTTTATGGAT ACACAACACA TATTTGCCAC CTGTCTCCCC 2700

ATTAAAATAT AATCTTCAGT AGAGAAACTC CATATCTTGT TAATACCTGA AACAAGAATA 2760

TCTTCAAAGA GTTCCTGGGA CATAAAAACG CTCAATTAAT ATTTATGTTA AACAGGGATC 2820

TGGGGTATAT CACAGAGGTA GAGGGCTTAC CTAGGAGGAG TTGGGCCATG GGTTCAACTT 2880

CCAGCACAGA ATGAAAGATT ATGTTAAATA AAGTTGGGAA GGATGTATGC CAGTCTATGA 2940

GTAGTATAGG AGGTAAATTA TGAATTCATA TTTACTTTTC GGACAAGAAG TGTTGTAGTC 3000

TTTATTTGAA ATAAAATACA TCTTAATTAC CAATAACAAT TGGTAAGGAG TGAATTCTCA 3060

AGCTGTGGCT TCCTGGTAGA TGAGTCCTGG GAGGTTTTCT ATTTCGATGA TGGTAGATAG 3120

GTAACCTGTC ATATACCACA TGAAATACCT GTGGCTTTGT AAACACACCG AGCAGTCAAG 3180

CAGGAGAATA GTTCCATACA GTTCGCGTCC CTTAGGATTG GTTTCGGGAT ACTTCTGGAG 3240

GTTCATTTAA ATAATTTTCC CCGAAGTACA TTATGGGCAG CCAGTGTTGT GATGGGAAGC 3300

TTCTGCCTGT TTTGCTTTGC GTCGTGCTCC ACACCTTTGA CAGATGTGCT CTCATCTGTT 3360

TACTTCTTTT TCTACACACA GAGCACAGCA TTAGCTGCTG TCCCGGCTTT GGATGTTATG 3420

TCAGCACATG AGGGACAGAC CTTCAGCTTA TCGAGTATTG CAGCTCTCTG TTTGTTCTGG 3480

AGCCTCTTCT GAGACTATGG ACTTAGTTCA AGGCCGGGTA ATGCTATTTT TTTCTTCTTT 3540

TTTCTAGTAG GAGGACAAAT AGTGTTTGCT TTGGTCACTC AAGTTCAAGT TATTGGATCA 3600

TGGTCCTGTG CACATATAAA GTCTAGTCAG ACCCACTGTT TCGGGACAGC CTTGCTTTGC 3660

TAGGCAAAGA GTCTCCCCTT TGGAAATTTT CCTGCTTTTG CAAAATGATG ACTATTTCTT 3720

TGATTTGGGG AATTGCCGTG TTGGTGAGCT GTTGCATATG GTTTATTGTT GGAATAAGGA 3780

GAAGGTATGG AAAGATTTTT AAAAATTTGT CTTTTAGCTT ATTTCTAGTA TTCATTGCCT 3840

TCACTATTAT GTAGTGCAAA AAATACTAAT GCATTAATAT TTTTAAATTT AAAATTTAAA 3900

GACGTACTTC TTTGACTAAA TCTAGTAAGA TGTAGAGAGT CCCCCTTGGA ACATTCACAT 3960

ATGCCACTGG TAATGCAGAT CTTGTGAAAT ATAACTAAAG AAATCACAAG TCATCGATGT 4020

AAGTTTGTGT CTGCATGGGC GGAACAAACC TAAGCTAAGA AGAGTAGTAT TTGGGAGGGA 4080

TCTTTCTGTG ACATGAACTG AATAGACGCA CTGCCTCAGC AAACACACAT TCATTTGAAT 4140

TTTCCTCAGA CTCAGTCTAA GCCTGGTGAG AGCACCAAGT GTGAGTCTGT CTGCCACTAA 4200

CGTTTCCTTC CAGTGGTAAT CAGCTGTGTG GCTGTGAAAC CTTGGCGCCT GCACATGACA 4260

GCCATTTGAA TAGTTCAAAG AACATTTAGG GACAGGATAT TAAGATATTT TCTGTGATGT 4320

CAACATCAAA ATAGGAGAAT GCCCCTGGCA TTATCTTCAG AGAGGTAGAC TACTGTGCGT 4380

TGTCTTACTT TAAAGAAATT TCTTTGCCCC TTTGGCTATT TTAATTCAAA CCTGAAAGTT 4440

TTCAGTTTTA ATTAAACTGT TGATTTTCAT GCTAGGAAAG GAAATATCAA TTATACTTAA 4500

TTGTTCTTAC AAGAAATAAA ATCATTTATG TCGGGAGATA AATAAGCTCA TAATTTTAAT 4560

AAAACATTTA AGAGAGAGAA AAAGAGTAGT GGATTATAGT TCATTGTCTG TCAATGTTTA 4620

CCTGACCCAG TTTCATTTTA TAATTATCTA ATTTTTCAAA TGAGATTCCT GTTCTTTCCA 4680

AATATCATTG CAGAATACTA ACATTCTTTT TTTCAGAGTT GAGAATCAAA TGGAGGGTTT 4740

TTTCATCCTG GCACAAGCTC CGCTCTTCAG TAACACCTCC AGCCCTCAGA ATGCCAATAT 4800

TTTAAATTAT GTAGGTTGTT AAAACTTTAG TGCTGGGGCT GGGGATTTAG CTCAGTGGTA 4860

GAGCACTTGC CTAGCAAGCG CAAGGCCCTG GGTTCGGTCC CCAGCTCTGA AAAAAAGAAA 4920

AAGAAAAAAA AAAACTTTAG TGCTGTAGCC CTTTCTGTTA TTTGATGTTT CACATCTGTT 4980

AAAAAACAAA ACAAAACAAA AAAAACAAGC AAATGGAACA TTTTAGGCAT TCTTTGGGGG 5040

AAATGATTCT TAGAGCAAGT CTAATCATTA GGTGATAGTT TCATTTTTAC ACCAAGAACA 5100

AGAATCTTGT TGGCTGTGTT AACACTTTAA GCCCTGTTGT AGGGAAAAAG CAATCAGACA 5160

CAGGCACAGA AAAGAATTTG GATGAGTACT TGATGATGTA TGTATATATG GTGAATAGAC 5220

TGATGGGTGG GCTGCTGGCT GGGTTGGTAA GTGGGTAGAT TTTTTTTTAA AGATTTATTC 5280

ATTTATTATA TATCAGTACA CTGTAGCTAT CTTCAGATAC ACCAGAAGGG CATCGGATCT 5340

CTTTACAGAT GGTTGTGAGC CACCATGTTT TCCTAACCTC TCAAGTCTCT GTCTTCCAGG 5400

AAAGCTGGTG AACCTCCTTT GGAGAACGGG TTGATTCCGT ACCTGGGCTG TGCTCTGAAA 5460

TTTGGATCTA ATCCTCTTGA GTTCCTAAGA GCTAATCAAA GGAAGCATGG TCACGTTTTT 5520

ACCTGCAAAC TGATGGGGAA ATATGTCCAT TTCATCACAA ACTCCCTGTC ATACCACAAA 5580

GTCTTATGTC ATGGAAAATA TTTTGACTGG AAAAAATTTC ATTACACTAC TTCTGCGAAG 5640

GTAATTAATT CGTTATACAG ATTCTGTTTG TTTCCTGGTC TGTTGATGTA TTAGTGTATT 5700

TAGTTGTTCC AATTTTGTTA GGTTGCAGAA TAGAGGTAAC ATAAAATCAG GGCGTTTCTT 5760

AGTAATAAGC ATTAGACATT TAAGGCAGAT GTAAACCTGT CATTGATGAT TCCGGAGACA 5820

GAGGACACTG CAGGAATCAG GAAGGTACAG ATTCATAGCA CCACTCGTCC CTTAACAACA 5880

CCCTGAGCAG GGTGTTGGCA CTCTTAGCCT TCAGTCCTTG TACACACGTT TCATTCCTAA 5940

GATATAGGCT GTATATTTAA ACACGATTTG GAAGCCATCA AGAATCTGTT CTAGAGAAAA 6000

CAGCATTTAA TGATCTTTTG CAAGAAAATA TCAGTTATAG TCTCTGTCAT TAAGTACATT 6060

GTAATCTGGT TAAAGAGTAT CTACTAAGAA AGTAAAGGCA GATTAGAACA ATACCAATGG 6120

ATGATGGGCC ATCCAGAGAA ATCCTACTGT AAATGCTGGG ATTTAAACTT GACCCCAAGG 6180

AAGAGTATGA CTTGATTCTA CCTTTGGAAT GTGCTGTAAA ATCATATTAG GGAAGGTTCC 6240

AGACAGAGAA GTGGGATGTA TTTAATCTAT CTTCCAGCCC ACTCTCTAAC ACTAGCTAGC 6300

TTTGGGCTTT AGACCCTCCC CATTTCATGG ATTCTATTTT CTACCAGGCA TTTGGACACA 6360

GAAGCATTGA CCCAAATGAT GGAAATACCA CGGAAAATAT AAACAACACT TTTACCAAAA 6420

CCCTCCAGGG AGATGCTCTG TGTTCACTTT CTGAAGCCAT GATGCAAAAC CTCCAATCTG 6480

TCATGAGACC TCCTGGCCTT CCTAAATCAA AGAGCAATGC CTGGGTCACG GAAGGGATGT 6540

ATGCCTTCTG TTACCGAGTG ATGTTTGAAG CCGGCTATCT AACACTGTTT GGCAGAGATA 6600

TTTCAAAGAC AGACACACAA AAAGCACTTA TTCTAAACAA CCTTGACAAC TTCAAACAAT 6660

TTGACCAAGT CTTTCCGGCA CTGGTGGCAG GCCTTCCTAT TCACTTGTTC AAGACCGCAC 6720

ATAAAGCTCG GGAAAAGCTG GCTGAGGGAT TGAAGCACAA GAACCTGTGT GTGAGGGACC 6780

AGGTCTCTGA ACTGATCCGT CTACGTATGT TTCTCAATGA CACGCTCTCC ACCTTTGACG 6840

ACATGGAGAA GGCCAAGACG CACCTCGCTA TCCTCTGGGC ATCTCAAGCA AACACCATTC 6900

CTGCAACCTT TTGGAGCTTA TTTCAAATGA TCAGGTAACT TTCCAGTGAC AGAAATTGCA 6960

TTTTAAACTC AAAACCCAAA AAGACTTATA GAGCTTTCTG TGCTATCAAC AAAGAAAGTA 7020

ATACTCAATG TCCGTGTTTA GCATGTGCGT AACAGAAGCA GCAATTTTTA GGTGCACAGT 7080

CCCATCGAAA GGGATGTCCC AGAAGCCACA GAACTCAGAC AGGTTGGTGC TCCATTAGTA 7140

CAGGTTCCCT GGCCTAGTCT TGCTCCTCAC CCGATATGTT CCTCTTAATA TCAAATTAAA 7200

TCCCCGAGTG CAGTCGTCAC CACCATATAA ACATTTGAAA TGATGACTGA CTTGCAGGTG 7260

TGATAAGAGC AGTGACCATA CCTTACTAAT TCACTGGAAT TCATAGGCAA AGTAACACCA 7320

TCGATTTTGT ATTCATATAG GAGCTGCAGC CATATTTTAA ATAGCACAAC TACTTGTTAG 7380

TCAAGCATTC TGAGGCTCAC TGTAATCAGG TAAAGTAGGT TTAACTCAGC GTCCTACCAG 7440

TTCCAGGCAT TGAAATGGAA TATCCTTTAT CCCACCCATT CAAAACGTAA TATATAAATG 7500

GAAGGCACAG TTTTGAAGGC CATGGTATGA TTTAGGGAAT TTACTCTCAT GGTCCAATCC 7560

CTTGTAATTG TATGCTAGGT GACATATCCT TCTGACTTAC TATGTTCATC GTATATTCAA 7620

TCCTTAGTTT ATAGAGACTG ACCAAAGCTC TGCTTTTGCA TAGCAAAGCT CCTTTTAATG 7680

CCCATTCCTA AACTCAAGGA CACGAATCCA GTTCAGTGCC CTTTTGCATA CTCCCTGGCA 7740

GACTCCCGTT GCCATACATC CTCCCTCGCT CGATTCCCAT GACCTCGCCC TTGCACACCC 7800

TGGTACTAGG ACCTCTCCTG GCGATACTTC CTACTACCTA TGCCACCTCA TTAAAAGGAA 7860

GGGATAATTG CTATTTACTT GCAGTTCTCT GAATGAGGAC ATTTTCCCCA TACGGCTCTT 7920

TCCACAGGAG TCCTGAAGCA ATGAAAGCAG CCTCTGAAGA AGTGAGTGGA GCTTTACAGA 7980

GTGCTGGCCA AGAGCTC 7997

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5537 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: YES

(vi) ORIGINAL SOURCE:
(A) ORGANISM: human

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 1..5537
(D) OTHER INFORMATION: /note= "Cholesterol 7α-Hydroxylase"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

TTTTTGGTTA TCTTTTCAGC CGTGCCCCAC TCTACTGGTA CCAGTTTACT GTATTAGTCG 60

ATTTTCATGC TGCTGATAAA GACATACCTG AAACTGGACA ATTTACAAAA GAAAGAGGTT 120

TATTGGACTT ACAATTCTAC ATCACTTGGG AGGCCTCACA ATCATGATGG AAGGAGAAAG 180

GCACATCTCA CATGGCAGCA GACAAGAAAA GAGCTTGTGC AGGGAAACTC CTCTTTTTAA 240

AACCATCAGA TCTCATGAAA TTTATTCATT ATCATGACAA TAGCACAGGA AAGAACTGCA 300

CCCATAATTC AGTCACCTCC TACCAGGTTC CTCCCACAAC ACGTGAGAAT TCAAGATGAG 360

ATTTGGATGG GGACACAGCC AAACCATGTC ACACTACCAT GCCTGACTTC CTTTCCATTT 420

TTGTATATTT GCTTGTTCTT CATTTGCCCG AGAAGTAACT CTAAAGGGCT GTATTATTTG 480

GATATTAGAT TGGCATTTTA TCTGACTGGG ATATCTTGCT GTGATTGTCC ATGTATAAGA 540

TCAGCTTTTC TATAAGCCAT ATTTTTAAAA AGATATATTA ATTTTTTAAA AATCCACCTG 600

TCTAAATAAA TGCACAAAGC CCCCCAAAAA CCTAGATTCT AAGAAAAATC TATGTACTGC 660

CATACAATGA TTGATATTAA TATTTATGGT GATAAATTAC ACACAAAAAA TGTGTGATCT 720

CTGTTTAAAC AGGCAAAAAC AAAAAACACA TGAAATAAAT CTATGGCATC TATAGCCAAA 780

ACTGGAAACA ACCCACATAT CCATCAATAG GAAATCAGTT AAATAAATTA TAGTACATTT 840

ATCCAATGGA AGATTAAGCA CATATTCAAT ATAATTATTT ATACACACAT ATAGATACAC 900

ACATGTATAA ATATAGAGAA TACTGTGGGT GTATGTGTGT GTGTGTTTAT ATACATATAT 960

ATACACACAC AGTACTGTTG CCTACCTTCT TTTGTCTTAA TTCTGTGAAC TCTCATTCAC 1020

TCTGCTTCAG TAGGATACCT CCTTCTTTTT GGTTCTTAGA CTCACCAAGT TGATCCTTGA 1080

CTCAAGACAT TGCATTTGCT GCTTCCTCTT CCTGGAATAT CCTTCCTTCT GATATTCACA 1140

TGAGTAGTCT CTTCTTGTCA TTCAGATCTC AAATGTCACA ATTTCAGAGA GCCCATCTCT 1200

GATCATCATA TCTAAAGTTG TCCTCATTCC CCCATAGCTT TCTATACCAT GTTTTATTTT 1260

TTTCATAACA TGTATTTTAT TACTCCTTTC TCCATTGGAA TAGAATCTCC ATTAGATTAG 1320

GAAATCTGCC TATCTTATTA ATGCCTGCAA CTGGAATACT TTTGAAGAGT TCTTGGCACG 1380

TAATAAATAC TCAACTAATA TTTTTGTGTA CACAGAAATA AAGTTTGGAA GAACAGATGC 1440

CAAATTGTTA CTAGTGGTTA CTTCTGAGTA AAGGAGTAGC ATGGTAGGTA AATTATTAAT 1500

AGATGTTCAC TTTCCACCAA GATATGTTTT AGTTAGTCTT AACTTACTTG AAATGAAATT 1560

TATTACTTTA ATAATTAGAA ACATTGATAA ACATTTTAGT CACAAGAATG ATAGATAAAA 1620

TTTTGATGCT TCCAATAAGT TATATTTATC TAGAGGATGC ACTTATGTAG AATACTCTCT 1680

TGAGGATGTT AGGTGAGTAA CATGTTACTA TATGTAGTAA AATATCTATG ATTTTATAAA 1740

AGCACTGAAA CATGAAGCAG CAGAAATGTT TTTCCCAGTT CTCTTTCCTC TGAACTTGAT 1800

CACCGTCTCT CTGGCAAAGC ACCTAAATTA ATTCTTCTTT AAAAGTTAAC AAGACCAAAT 1860

TATAAGCTTG ATGAATAACT CATTCTTATC TTTCTTTAAA TGATTATAGT TTATGTATTT 1920

ATTAGCTATG CCCATCTTAA ACAGGTTTAT TTGTTCTTTT TACACATACC AAACTCTTAA 1980

TATTAGCTGT TGTCCCCAGG TCCGAATGTT AAGTCAACAT ATATTTGAGA GACCTTCAAC 2040

TTATCAAGTA TTGCAGGTCT CTGATTGCTT TGGAACCACT TCTGATACCT GTGGACTTAG 2100

TTCAAGGCCA GTTACTACCA CTTTTTTTTT TCTAATAGAA TGAACAAATG GCTAATTGTT 2160

TGCTTTGTCA ACCAAGCTCA AGTTAATGGA TCTGGATACT ATGTATATAA AAAGCCTAGC 2220

TTGAGTCTCT TTTCAGTGGC ATCCTTCCCT TTCTAATCAG AGATTTTCTT CCTCAGAGAT 2280

TTTGGCCTAG ATTTGCAAAA TGATGACCAC ATCTTTGATT TGGGGGATTG CTATAGCAGC 2340

ATGCTGTTGT CTATGGCTTA TTCTTGGAAT TAGGAGAAGG TAAGTAATGT TTTATCTTTA 2400

AATTGCTCTT TGATTCATCC ATTTAATTTT TTTACCTTCA TTTTTATACA GTAAATTTGG 2460

TTTTCTATAC TTACACATAT TAGCATTATC TTCCTTATGT TTTAAATGAA AAATTTGATT 2520

TGAATTTTTA AAGTAATATC TTTTTTACTA TATCTCACAA GACATATGAC AGCTTCCCTT 2580

TTTAGTATTG GCATATACCG ATGGTAATAT ATAAATGTAT ATTGGTGTTA AACATAACTG 2640

ACAGAAATTG TATAAGGTCT CTATGTACAT TTATATGTGT ATCTAAAGAG GAAGCCCAGA 2700

TTAGTAAGGA TACAAGTAGC AAGTGGGAAT CTACAATGGA AAGGATTGCT TTCTCTCACA 2760

TGGCTTCAAT AGATACTCTT GCTTAAATAA ATGTTCTCTT TTAAGCTCAT TCTTGTGCAT 2820

CGCATAGACT CAGCCTAAGC CTGAACAAGA GCATAGAGCC TGAGCTGATC ATTCTATTAC 2880

TGTTTTTAAA TAAATGTTAA TCAACTGTGG TGAATTGGGA AAGTTTGCTG AGTGTATGTG 2940

ACATCGATTT CATTTATTTA CAACTGGTTC AAGAATGCAA GAAAAACAAA TACAGTCAGA 3000

TCCAGAACCA TAGTTTATTT AACTTCTAAT TGGCTCAAGG AGTAATTGTG GGGAGGCATA 3060

TAGATATTCT CTGCTATGTC AATCTCAAAA AGAGAAAATA ACCCTAACCA TCTTTCAGCT 3120

TTGTAGATTG CTATGTGTTT TCTGCCTTTG CAGTTTCTTT CAGGCCTGAT AGTTTTTACT 3180

TTTAATTAAA CTACTTATCT TCAAACTAAG AAAAGAAAGG TAATTACTTT ATACTGTATT 3240

ATTCTATCAA GAGGTACAGA AGTTTATGTT GGAAAATAAG TTTACATGTT CTAATAAAAA 3300

CATTTTAAAG GAGCACTGAA TTACAATAGA TGATTCCGTC AGTGTTTATC TTACTCAATT 3360

TCATTTTATA ATAAGCTGAT TTCTCACATG AGATTCTTCT TCTCTGAAAC CATCCTTATA 3420

GAATATAATA TAGATATCTT TAAACTAGGA ATATTTTCAA AACCTCAGTT CTGAAATCCT 3480

CCCTTATTCA GTGATCTGTG TCTTTAAAGA AAATAATCAA AAGAAACATT TTGAGATATT 3540

TAGAAAAATG ATGCTTAGCA AAGTGATAAA CACTAGAATG TAGTTTTGTT TCCGCACTGA 3600

CAACAAGAAT CTTGTTGGTC TTGTAAATCC TTTTGCCTGT ATCACTGGGA AAAGTGATGA 3660

GCACATAGTA GACGGGTGCT TGTTGAATGT GTATATGGAC GGATGCATGA ATGGATGGAT 3720

TTAGTAATCC TTTCCACCAA CATATCATGT TACTAGGTTA ATATAACCTA TTACTGTAGT 3780

AAAAGAGCAG GGCCCATCCA ACAAAAGAAA TATCTATAAA CTATAGGGTT TCAAAGTTTG 3840

AAGTCAGTGG GAAAAATTTT AAAACCTGAT GTAAGTAAAA ACCCAAAACT GTAATCATCC 3900

ATGTCTATCA TACACTTGTG TCTGACAGGC AAACGGGTGA ACCACCTCTA GAGAATGGAT 3960

TAATTCCATA CCTGGGCTGT GCTCTGCAAT TTGGTGCCAA TCCTCTTGAG TTCCTCAGAG 4020

CAAATCAAAG GAAACATGGT CATGTTTTTA CCTGCAAACT AATGGGAAAA TATGTCCATT 4080

TCATCACAAA TCCCTTGTCA TACCATAAGG TGTTGTGCCA CGGAAAATAT TTTGATTGGA 4140

AAAAATTTCA CTTTGCTACT TCTGCGAAGG TAAGCAGTTT TACATTTATA TACCATTCTG 4200

TTTGTCTTCT ACCTTTTTAT GTGCTTGTCT ATTTAGAAAT TTTGATGTAC TTAGATTTTA 4260

TGATAAAGGT GTTGAAGAGA GTTATCCTTA TGTGGAGATT CTTAGAAACA TAAATAAATT 4320

ATACGTAGCT TCTTAGTAAT AATCATTTAG AAAGTCAAAA TAGGTATAGA TTTCCGTCAT 4380

TTGCTTTGCA CGAGCTAATG AGGGTGAAAT ACAGATTAAA TGCTCTACTG AGACAGGTGG 4440

CACTGTACGA ATAAGATAGA TTAAAATTCA TCACATCAGC AATGTCTATG CAGAGCGAAG 4500

TGACGGAAAC CTAACATTCA GCAGTTGTCT CACCACACTT GTGCCACACA GTGTTTCATT 4560

TTGATAAGGA ATTGGCAAGA TATTTTAACA TCATTTAGAT GTAATAAAAG AAGATCTGTT 4620

ACTGAGAAAA AAAACCAATA ACTACTTACT TACTGCAAAT AAATATTAGC TTTGGTCTTT 4680

GTGACTAAGT AGCTTAAAGT TTGGTTAAAA TACATCTACA GCTGGACACA ATGGAACACA 4740

CCTGTAGTCC CTGCTATTTG AGAGGCTGAG GCAGGAGGAT CGCTTGAGTC CAGGAGTTTG 4800

AGGCTGCAGT GAGCTATCAT TGTGTCACTG CACTCCAGCC TGGGTGACAA TGTGAGACCC 4860

CATCTCTAAA AGAAAAAGAA AAAGAAATCT ACAAATAATA TAAAAGATAA CTAATGATTT 4920

TAAAACATTA TCAATTAGTT TATGTGCAAT AGCTGTAAAT AAGTGCAGTA GCATAAGAAA 4980

TAAGACATAG ATGACTTGAG TGATCCAGGG GAGTGCCACT GAAGTTGGCT TTAAAGGAAA 5040

GGTACAGTTT GGTCATTTAT TTGTAAAGTG CTATGAACTT GTACAAGGGA AAGCCAATTT 5100

CCCGTGTTTA CCAAGTAAGG AACTATGAAA GTATCTAATC CGTTTTTCAG TCATTTACTA 5160

TGACTAGGTC AGGTTTAACT TCTTTTTCTG CATGTTTTAT TTGCTATCAG GCATTTGGGC 5220

ACAGAAGCAT TGACCCGATG GATGGAAATA CCACTGAAAA CATAAACGAC ACTTTCATCA 5280

AAACCCTGCA GGGCCATGCC TTGAATTCCC TCACGGAAAG CATGATGGAA AACCTCCAAC 5340

GTATCATGAG ACCTCCAGTC TCCTCTAACT CAAAGACCGC TGCCTGGGTG ACAGAAGGGA 5400

TGTATTCTTT CTGCTACCGA GTGATGTTTG AAGCTGGGTA TTTAACTATC TTTGGCAGAG 5460

ATCTTACAAG GCGGGACACA CAGAAAGCAC ATATTCTAAA CAATCTTGAC AACTTCAAGC 5520

AATTCGACAA AGTCTTT 5537

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2575 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: YES

(vi) ORIGINAL SOURCE:
(A) ORGANISM: human

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 1..2575
(D) OTHER INFORMATION: /note= "Cholesterol 7α-Hydroxylase"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

GAATTCTACT CTTTAAAGGG GTGAATATTA TGGTACTTGA ATTTTATCTC AAGAAAAATG 60

AATAAAAAGT AACTAAATCA TTGAAAATAT CTGATGGCAT GGGGTTTGTG GGGTAACTGG 120

CATTCCACAG TGATTTTCAA AGGGCTTGTG CTGTTTTCAT TTTGCTTTGT TTTAGTTATG 180

GAGCCCTTCC TTGAAACAAA CTTCATACTA CAGTCCTCTT TCATGAAGCA GAAGAGGGCA 240

GTGGGCAGAG CTCTCCTTTG GCTTTCTCCC CCACCACAAC AGGGAGCCCT GGAGCTCTAG 300

GAGAGAAAAT CTGAAATATA AAGGGCATGC ATGTGAGCTG TGGAGTCCCA GAGCCCTGGG 360

TTTGCATCCT AGATCTGCAA CTCCCGTGAA TTGAGTTTTG GGAAGTTGCT GAAACTCTGA 420

CCTCCTGTTT TCTCATGGTA TTGTTGTAAG GGTTAAATGA GACAATGTAT GTGAAGACCC 480

TGGCCCCACA GTAGAGGCTC TGCACACATT TCAGCGATAC TTTCCTCATG TATTTCCAAA 540

AATGTTTTCT CATTTTCTTA AAATGTCAGA AAGAAGACAA CAGAACTTAC TTGCCTTTTA 600

CAACAGAACA AATGGAGCAA GTCAGAGGTC AAGGTGCTAA CATTCTTCAT GGTTCCTCAC 660

CACCTTTTGT TCTGTTAGCC TATAGGGAAA AGTCTTCTTT CTCATCTCAT TATCTGCAGG 720

GGAAAATAGT ACTTCAGCAA GTGATCCAGT TGAAGAACAT CTCCAGGGCC ATTAACATAC 780

AGAGGTTTGT TCTACTCTCT CTGTGCTCCA TGTCTAAGAA CCTCAGCCTT CCTCCTAGGA 840

GCTAGGGAAA GTCAGGAAAG TGAAAATAGT ACCCCAGCTA ATGAACTGCC CTGTGCTGGC 900

CTGAGAAGAC AAGACCAGCT TCCTCAATGG CTCAAGATTT GGTTTCCTTC AATATGTCCT 960

TTTGGAAATA TGTCCATGAC ATCGGAGAGA TAAAAGGAGC CAGGATTGCT CACATTCAGG 1020

AAAAAAGCTC CACTATCTTT CTCTCTCTCC CTCTTTCTCT CCCTCCCCCT GACTGCCCTC 1080

TTCTCTATCT CTCTCTCTCC CTGAGCTGGC AAGGTTAATT GGTCGCAGAA AGCCGAAGAA 1140

ACAAGTGGGC CTCCTGGAAC AAAGTTCAAA AAGCCGAAAA CGGGAAGAAA ACTAACCACA 1200

AAAGTAAAGG AACCACTTAG CCTTCTTTGA TTCCAGGCCC CCAAGCCTGT CTTTAACTTG 1260

GATGAATGGA GTTCTTCCTG TGCTACAGCA CCGCATAGTA GGGGCTGCCC TGGGCCTGAA 1320

GCCAGAGCTT CACCATATTC AGTCATCTGT ACATTGAGGC AACAGTGCCT GCTTCATGGT 1380

GCTACCCTGT GGATTAAATG AAGCAAGTTT TTGATGATCT TGACACTGAA TATTGATGCA 1440

TTGGTCAGAC TTTTTCTGAT AGTAAAAAAT GGTGGTTTCT TGTTGTCAGA AATCAAATCA 1500

ATATATTTGT TCTCCTGTTG ATTAGCTATG TCCCCTAGAG GGCAGCGACT TTGCCTGTCT 1560

TATTTATCTC TGCATCTCCA GCACTTAAAA GGTGCCTTGC ATAAGGTACA TATTAAGTTC 1620

ATATGAATGA ATGAATGAAA TGCATATGAT TTATTCATAC CCAGTTGGTG GTGTGTTTAC 1680

CCTTTCCTAA ACCTGTAGTC AGATGGCCTT TGAATCCCCT GTACTTCTTG TGAGGTACTG 1740

TGCTGTAAAG GTGGACTATC ACACTTCAGT TCAGAGCAAT CTGGGCTTGA ATCCTGGATT 1800

TGCCAGTTTA TTAACTATAG CAAACATTTT TGAGCATACA TTGTGCCAAG TGCTAGGCTA 1860

ACTGTCTTAC ACACATTGTC TTATTTCGTC TTAATATCTA TGAGTCATGC ACTATAATCA 1920

TCCCCATTTT ACAGATAAGA AAGCAAAGAC TTGGAGAGGA AAAGCATCTT GTTCAAAGGT 1980

AAATACTTAA TGGCCAAGCC AACATGCAAA TCTAGATTTA ATTGCAGCTT CCTCTTCATC 2040

TACCATTCGA ACTAATTCAA GCTATGTAAT ATTTCCCACT GAACCTTCTT GCCTCTACTT 2100

CCTCATCTTT AACATGGTCA AAATACCTGT CCTGCCCAAG TTAGTTATTT CATTAAAGTA 2160

GAAAAATACA AGAGAAGCTT TTAAAATGTG AAACCTCAAA TGAATGTAAA ATTATGATGA 2220

TTCCTTTAGA ATTTGTCAAC ACCTTCTTTT CTCTACTCCT GCTAGGCATT TACAATCTCA 2280

AAACCATGTA TTTAAGATGC AAAACTATAT TTGTATTTGC CATAACTGGT TTCTTTCCCT 2340

ATGGCTTCAT GAAAATGTGG CTCGAATGTG TTTATTATGA AAGCCCCAAA TTAATCACGA 2400

CAAGACTTCA CCAGCCCATT CCACAATAGA CTCCCATTAC TTTGCCCTGA CTTAGAAACC 2460

TCATATACAG TCTTGATTCA GTACAGCTCT GTGATGCTCT TGGAAAATGC AAAGTGCTTT 2520

CTTAATTGAG GCAATCTGTG TCCCACTACA GAGAGGTGGT TTAACTTGTG AATTC 2575

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2316 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: YES

(vi) ORIGINAL SOURCE:
(A) ORGANISM: human

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 1..2316
(D) OTHER INFORMATION: /note= "Cholesterol 7α-Hydroxylase"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

AGAGCAACCT GGGCAACATA GCAAAACCCT GTCTCTGCAA ACAATAAAAA GAAGAAAATT 60

AGCTGGGTAT GGTGGCACAT GCTATAGTCG CAGCTACTCG AGAGGTTGAG GTGGGAGGAT 120

CAGTTCAGCC TGGGAGGTTG AGGCTGCAGT GAGCCAGATC ATGCCACTGC ACTGCAGCAT 180

GGGCAACAGA ATGAGACCCT GGCTAAAAGA AAACAAAATA AAAAATTCAG ACACAGGTTG 240

AATCATTGAT AACAGCATAG TGGTAACAGA AAGAAAGTTT GGGAAATTTT TATCTGATCA 300

GCTTCCCATA CCCTGTTCAT CTTTGTGTTA TGCACTGCCA GGCTGTCTGT AGGTTCAGAC 360

TCTATATCAT ATGACCTTCA AACACTTGGT TTGTTCTTCT CCTTCCTTCC TCCCTTCTTC 420

TTTCATTTTT TATCTTTTTT TCTTTTAAAA TGTTTAGATA GTATAATAAG GAACTGCTGA 480

GGCTTTCCAG TGCCTCCCTC AACATCCGGA CAGCTAAGGA GGATTTCACT TTGCACCTTG 540

AGGACGGTTC CTACAACATC CGAAAAGATG ACATCATAGC TCTTTACCCA CAGTTAATGC 600

ACTTAGATCC AGAAATCTAC CCAGACCCTT TGGTAAAGTC GCAGTGTGCC CGAATTGAAA 660

TTCAATATCC AGGTGATAGC TACCTAGATC TAAATAAAGA GGAAATTTAC AATGGTAGAA 720

TTGATTTTCT CATAGTAGTC ACAGGAATTG TCTGACTTAA TTGTGTTAAA TATTCATATA 780

TTTTGGAAAA TTTAGATAGT GGTCTGAATT TTTCATTTTA GTCCTGATAT TTGCCATCAC 840

ACAGTCTTTG CTAGATTATA TTTGCAGTCA TGATAATAAA CCTGCCACTT TTTTTTCTT 900

AAAAAGCACC TCCTCCCAAA TCCAGGAAAT TGGAGGCTAA TATATTGATT ATTCTAGTTT 960

CTTCTGGGAA CCCTTCTCTC TCTAGCTCTG CCTGACTAAG GAACTAATCG TTCAAGCAGG 1020

ATAGGAAGGT ATCACAAGGC TTCCTTAGCT GCATTAAGCT CCTGTTCCTT ATTACTTTCT 1080

GATTCAATGT GGAGTATTTG CTAAATCACT AATGGGGTAG AATTAAAAAG AAAATTACTC 1140

TTTGGAGCTT CCAGGTTTAG AAAGAGATAA ATTTCTTTAA AACTAGCTTA AAGGCGGTTT 1200

TCTTTGTATT TTTATTGCAG ACTTTTAAAT ATGATAGGTA TCTTGATGAA AACGGGAAGA 1260

CAAAGACTAC CTTCTATTGT AATGGACTCA AGTTAAAGTA TTACTACATG CCCTTTGGAT 1320

CGGGAGCTAC AATATGTCCT GGAAGATTGT TCGCTATCCA CGAAATCAAG CAATTTTTGA 1380

TTCTGATGCT TTCTTATTTT GAATTGGAGC TTATAGAGGG CCAAGCTAAA TGTCCACCTT 1440

TGGACCAGTC CCGGGCAGGC TTGGGCATTT TGCCGCCATT GAATGATATT GAATTTAAAT 1500

ATAAATTCAA GCATTTGTGA ATACATGGCT GGAATAAGAG GACACTAGAT ATTACAGGAC 1560

TGCAGAACAC CCTCACCACA CAGTCCCTTT GGACAAATGC ATTTAGTGGT GGCACCACAC 1620

AGTCCCTTTG GACAAATGCA TTTAGTGGTG GTAGAAATGA TTCACCAGGT CCAATGTTGT 1680

TCACCAGTGC TTGCTTGTGA AATCTTAACA TTTTGGTGAC AGTTTCCAGA TGCTATCACA 1740

GACTCTGCTA GTGAAAAGAA CTAGTTTCTA GGAGCACAAT AATTTGTTTT CATTTGTATA 1800

AGTCCATGAA TGTTCATATA GCCAGGGATT GAAGTTTATT ATTTTCAAAG GAAAACACCT 1860

TTATTTTATT TTTTTTCAAA ATGAAGATAC ACATTACAGC CAGGTGTGGT AGCAGGCACC 1920

TGTAGTCTTA GCTACTCGAG AGGCCAAAGA AGGAGGATGC TTGAGCCCAG GAGTTCAAGA 1980

CCAGCCTGGA CAGCTTAGTG AGATCCCGTC TCCAAAGAAA AGATATGTAT TCTAATTGGC 2040

AGATTGTTTT TTCCTAAGGA AACTGCTTTA TTTTTATAAA ACTGCCTGAC AATTATGAAA 2100

AAATGTTCAA ATTCACGTTC TAGTGAAACT GCATTATTTG TTGACTAGAT GGTGGGGTTC 2160

TTCGGGTGTG ATCATATATC ATAAAGGATA TTTCAAATGT TATGATTAGT TATGTCTTTT 2220

AATAAAAAGG AAATATTTTT CAACTTCTTC TATATCCAAA ATTCAGGGCT TTAAACATGA 2280

TTATCTTGAT TTCCCAAAAA CACTAAAGGT GGTTTT 2316

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10614 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: YES

(vi) ORIGINAL SOURCE:
(A) ORGANISM: hamster

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 1..10614
(D) OTHER INFORMATION: /note= "Cholesterol 7α-Hydroxylase"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

GAATTCTAAA CACATATTAA TATCAATGAC TTATATGTAT GTATATATAT ATCTAATATA 60

GATAATGTAT CTAGGGATAT ATATATATGT ATATTTTATC TTTCTTCCTT TTATTCTTTC 120

TTCTCCCCTC TCTGTTCAAC ACCGAGGAAT AGAATGCACT GTGGTGTCAT ACTCTGCTTA 180

CTCAGCCTCT TATTGACCTC TGAGTCAATA CAGTGCTGAT GTACATCTCC AAATGCCCTC 240

TTTTCTCCTA ACCACAGACT TTTACATTCA GTAATCAATT TGACATTGTC CCATGATTTA 300

CAAATGTTCA CAATAGTATA TTGACCTATT GCTGCCTTCC AAGGTCCTCT CCCACTCCCA 360

AACATCCCAA TATGAACCAG CTTTTGCCTA TCTTCTTGTC TCTTACTTTA ACTCAATGTC 420

ATTCCCTATT CACTTTGCTG TAATAGATGC TACCTTGATT CTGGTTTTTA GCACCTTAAT 480

TTCGCTCTCT GCTCAGGAAC TCTGCCTTTG CTGTTCCCTC TTCTGGGAAC GCTTTTCCTT 540

TGCTGTTATA TCTCTTCAAA ACAGCTTCTC TATTCAATAT GCTCAAGCTG CCTTCAGCCC 600

TCAACAGCTC TCCCTACCTC ATTCTAGTCC CTCCACTAGA ATAGAATCTT CATGAGAGTA 660

GCGAACTTCC CTATCTTGCT AGTACCCAAA GGCAGAAAAA TCTTTAAAGA GTTCCTGGGA 720

CATAGAAAAA GTGCTCAATT AATATTTGTA TTAAATAGGG ACCTCAGGTG TAACTCCGTG 780

GTAGAGCGTT TGCCTTAGAG AAGTAGGGCC ATGGGTTCAA ATTCCAGCAC AGAACAAAAA 840

ATTGTGCTGA ATAAAGTTTG GGAGGATGTG TAGCAGTTTA TAGTGCAAGT GGCATAAGCA 900

GTAAATAATG AATTTGTATC CACTTTTCTA GCAAGAAGTA TTTTATTCTT TATTTGAAGG 960

ATAACAATTG GTAAAGACTG CATTCTCAAA ATAAACTATG GCTTATGGCT ACGTGGAAGA 1020

TGAGATAGGG AGAAGGTTTT TTTTTGATGA TGGCAAAATA ACATGTCATA GTCCACACGA 1080

AACACCTGTG AAGTTGTAAA CACACCTAGC AATCAAACAA GAAAATTGTC CCACCCTATT 1140

ATCATTCTTT TGGATTGGTT GTGGCATATT TCTGGAAAAT GATTTAAATT AATTCCTTCT 1200

AAAGGTAACA ACACAAACAA CCACTATCAT GACGAAAAGC TTCTGCCTGT TTCAGTTTAC 1260

ATCATGCTCA ATGTCTACAA CAGACGTGCT CATCTTCAGA GTGTTTACCT CTGCTTTTTA 1320

CACACATTGA AGCACAATGT GAGCTGCTGT CCCTGGGTCT GAATGTTATG TCAGCACACA 1380

AGGGACAGAG CTTCGGCTTA TCAAGTATTG AAGCTCTCTG CTTGTTTTGG AGCCTCTTCT 1440

GATACTATGG ACTTAGTTCA AGGCTGGGCA ATACTATTTT TTTCTTTTTT CTAATAGGAG 1500

GACAAATAGT TAGTTGTTTG CTTTGGTCAT CCAAGTTCAA GTTATTGGAT CATGGTCCTA 1560

TGTGTATAAA GAGTCTAGTT TGAGCCTTTC AGGGGCAGCC TTGCTGGCTA AGCACAGACT 1620

CTCCTCTTGG GAGTTTTCCT GCTTTGCAAA ATGATGACCA TCTCTTTGAT TTGGGGGATT 1680

GCTATGGTAG TGTGCTGTTG TATATGGGTT ATCTTTGACA GAAGGAGAAG GTATGTCTTT 1740

TAGCTTATTT CTAGTGTTTT CACTATTATA CAGTTCCAAA AAAATACTAG TACATTAGTA 1800

TTTTTATTTA AAATTTAAAG CCATGCTTCT TTGACTAAAC CTGACAAGAT GTAGAGTTTC 1860

CCTTTGAATA TCCACATACA CTGATGGTAA TGCTGATCTT GTTAAACATA ACTAAAAAAA 1920

TTATAAGTAT TGATGCATGT TTGTGTGCAC TTCTGTGGAG TACACCTAAG CTGGGAAGGG 1980

TGCATTTGGC AAGGGTGACG TTTGGAAAGG ATCTTTCTCT CACAATAACT GGTTATGCAT 2040

ATGCTCTTCT GGGTTCTCTG TTACATCAAC ATTAAAATAC AGGAATACCC TTGGCATATC 2100

TTTGGCAAGG TAGACTGTGT CTGCTGTCTT AGTTTTAATA ACTTCTTTGC CTTTTGAGTT 2160

ATTTGAATTT ATGCCTGATC GTTTCCAGTT TTAGTTGTCT TAATGCTAAG AAAGGACAAA 2220

TCAATTATAT TTAGTTATTC TAACAAGAGA TAACTAGTTT ACGTTGAAAA ATAAATTATC 2280

TTATAATTTC TAATAAAAAC ATTTAAGAGA GTTAGAAATC AGCGAATTAT AGCTGATGAT 2340

CTGCCAATGT TTACCTCACT CAACTTCATT TTAGATACTT TTTCAAGTGG GATTCCTATT 2400

CTCTTCAAAT ATCCGCACAG AATTATAGTC CCCTTCTTTC AGAGTGGGGG GAATCAAATG 2460

AAAGGTTTCA TGTGTGCTAG GCAAGAGCAC CACCGTTGAG CCACACCTCC AGACCCCACA 2520

ATGCCAACAT TTTTAAACTA TGTAGAGTTT AAAAAACTTT AGTTCTGTAG CCTTTTCTAT 2580

TAGCTGGTGT TTCATGTCTT CAAAGAAAAG GAAAACTGAA ACATTTTAGA CATATGGACA 2640

AATGATTCCT TGAACAAGTC TAAGCACTGA TGATAGCTTC TTTTCTACAG TGAGATCAAG 2700

AATCTTGTTA GCCCTGTTGA TACTTGTAGC CCTGTCACTT GGAAAAGCAA TCAATTTTAT 2760

GATCTAGAAA ATAGAGCTTG CCTAAAGATC AGAGTGCAGA GCTAGTCACA CTAGTCAGCC 2820

ATACAGGTTA GGCAGTGGTG GCACATACCT TTAATCCCTG CAGCCACTCA AGTTACCCAT 2880

AGAAGCTGGG TGGTGGTGGT GCACACCCTT AATATAAGGT GGAGCACACT TTAATGTAAG 2940

GTGGGTAGAG TCAGGAGTGC AGTGTATTCA GTCTGCAGTC ACACTGAGAA CAATATCACC 3000

CCAGTCTTGT TAGAGGTAAG AACTCTCTAG TGATTGGCTG CTTTGCTCTT CTGATCTTCA 3060

GTTTGAACTT CTGTCTCTGG GTTTTTATTA TTCGTGCTGC AGACATAGAC ATAGCAAACA 3120

ATTTAATGAG TGATTGATGA ATGTAGATAT GTATGTACAT ATTGTGCTGG ATAGACTGTA 3180

GATGGGTTGG TGGATGGGTT GATGAGTGGG TAGATTTAGT AATCACCTTC ACCAATATCT 3240

TAGTAGGCTA AAAAGCCCAC TGTTTTAGTA AAAGAGTGGG GTATCCAACA AAGAAGTATC 3300

TATAAACTGT AGTTATGTGG TAGAAATAAG GGGTAGAAAC CAGTAAAAAT TCGGCTTATG 3360

TACAAATGCT AAACATGTAA TTTCCTAAAC CTCTCAATCT GTCTCACAGG AAAGCAGGTG 3420

AACCTCCTTT GGAGAATGGG TTGATTCCAT ACCTGGGCTG TGCTCTGAAA TTTGGCTCTA 3480

ATCCTCTTGA GTTCCTGAGA GCAAATCAAA GAAAGCACGG TCATGTTTTT ACCTGCAAAT 3540

TAATGGGGAA ATATGTTCAC TTCATCACAA ACTCCTTGTC ATACCATAAG GTGTTATGTC 3600

ATGGAAAATA CTTTGATTGG AAAAAATTTC ATTACACTAC TTCTGCAAAG GTAACTAGTT 3660

TTTACAGATT TTGCTTGTTT ACTAGCCTGT TTATTTATTA GTTTATTTAG TTGTTCCAAT 3720

GTTATTAGAT TGTAGGATAA AGGGAACATA AAATCAGGAA GTCTCTTGGT ACTAAGCATT 3780

AAAAAGTCAA GGTAAATGTG AATTTGTGAT TGATGATGAC ATACACAAAT TAAGCACTTT 3840

GTAAGTACTT TCTGAGCCAG AAGACACTAC AGGAAGGCAC AGACTCATAA CATCCATGCT 3900

GCCATCTACA CAACACTCAG AGCACTCAAT TACCACATCA TGCACACGAA CTCGTTCGTT 3960

AAGAAGTCGA CAGTATATTT AAGCATCATT CAGATGTTAT CAAGAATCTC TATTCTAGAG 4020

AAAACAACAC TTAGCTGAAT TTTTACAAGA AAATATTAGA CATGGTCTCT GTCTTAAGTA 4080

GATTAAAGTC TGGCTAAAGT GCATCTGCAG AGAACAAAAG GTAAAGATAA AATCAATGGC 4140

CCATTAGTCC AGAGAAGCTT ACCTGAAAAT CTGGGATTTA AACTTGACCT TAAAGGAAGA 4200

GTATGTCTTA AGTTTGACTT TGAAAAATGT TATGAAATTG TATTGGGAAG GCTAGACAGA 4260

GAAGTATGAT ATACTTTAAT CCATCTTCCA GCCATTTCCT AACACCCAGG TTTAGCTGCT 4320

CCCCCTCTGA CGAATTTCAT TTTCTACCAG GCATTTGGAC ACAGAAGCAT TGACCCAAAT 4380

GATGGAAATA CCACAGAAAA CATAAACAAC ACTTTTACCA AGACCCTCCA GGGAGATGCT 4440

TTGCATTCAC TCTCTGAAGC CATGATGCAA AACCTTCAAT TTGTTCTGAG GCCTCCTGAT 4500

CTTCCTAAAT CAAAGAGTGA TGCCTGGGTC ACCGAAGGGA TGTATGCCTT CTGCTACCGA 4560

GTGATGTTTG AAGCTGGATA TCTAACTCTG TTTGGCAGGG ATACTTCAAA GCCAGACACA 4620

CAAAGAGTGC TTATCCTGAA CAACCTTAAC AGCTTCAAGC AATTTGATCA AGTCTTTCCG 4680

GCGTTGGTGG CAGGCCTCCC TATTCACTTG TTCAAGGCGG CACATAAGGC CCGGGAACAG 4740

CTGGCTGAGG GCTTGAAGCA TGAGAACCTC TCTGTGAGGG ACCAGGTCTC GGAACTGATA 4800

CGTCTACGCA TGTTTCTCAA TGACACTCTC TCTACCTTTG ATGACATGGA GAAGGCCAAG 4860

ACACACCTCG CTATCCTCTG GGCCTCTCAG GCAAACACTA TTCCTGCAAC CTTCTGGAGC 4920

TTATTTCAAA TGATCAGGTG GATAGCAATT TGAGTGTTTA TTCTTCATAG TGACAGAAAT 4980

TAACAATTTT TAATAAACCC CCCAAAAGAC TAGCAGAGCT TTCTTTGCTG TTGGTCAAGA 5040

ATGTGATACT CAGTGCCTGT GTTTGACATA TATATATAAC AAAAGTAGCA TTTTGTAAGA 5100

ATATAGTCTC ACCAGAAAGG GATGTCCCAG AAGCCGCAGA ACTTAGATCT GCTGGCACTT 5160

GTCATTAAAG GTCCCCTTGC CCAGTCTTGC TTTTAACTCC ATAGTGTTCT TCTTAGTGTC 5220

AAGTTAAATC TATGACTGCA GTCTTCATCA CAACTTTAAA TAATGACTGA CTTGTCAATG 5280

TGGTAAGTGC AGAGGCCACA CCTTACTAGT TTGAACATTC CTGTTTTCTG CGGCCTCACA 5340

GATTTACAGC AGAGTTGCAA CATCAATTTC ATATTACCTA TGAACTACAA CCATATTTTA 5400

AGTTCAACAA CTACTTGTTA GTAACATTTC TGAGGCTCAG TTCACTTTAA CCAGATAAAG 5460

GAGATTTCAA ACAGCTGCCA ACAAATTTCC ATGCACTGAA TGGAAGTATT CTTTATCGCA 5520

CAGTTCAAAA ATAATAACAT AAATATTCTG AAGCTGTGGT ATGAATTTAA AGAGTAAATT 5580

TGAATTTCTA CTTGGGAATT CACCAATACC CTGTAATTGT ATGTTAGAGG AAGTATTCGG 5640

AATGAATTAC TCTACTCATC ACACGAATGT CTAGCCCTTA TTAGAATCAT TGGTTTATAG 5700

AGATCTGACC AAAGCTTTGC TTTTACATAG CAACGCCCCT TTAATGCTTC TTCATAAATT 5760

CAAGGACATG AATCCAGTTC AGAATACAGT ACAAGTAAAT GACAATGCCC TTTGCATGTT 5820

CCTGGAACCA CTTCCCTTTT CATGCTCCCA TGCTAACGCG ATCACCTCAT TAAAAGAAAT 5880

GGAGTTCTTA TTTACTTGCA GCTCTCTGAA TAAGGCAATA TCTTCCATAT GTCTCTTTTC 5940

ATAGGAGTCC TGACGCATTG AGAGCAGCCT CTGAAGAAGT GAATGGAGCA TTACAGAGTG 6000

CTGGTCAAAA GCTCAGCTCT GAAGGGAATG CAATTTATTT GGATCAAATA CAACTGAACA 6060

ACCTGCCAGT ACTAGGTGTG TTCCCTATGC TATCCCTCAC TAACATGTCA CTAGTAACAA 6120

TGCTCAACAT ATAATGAATG TACTATATTC TTGATATTTT TGCAACGCTG CAACAGTCTA 6180

ATAACTAGGG TCATCTTCAT TTTTTCTAAC AAACAAGGAA CTGAGACCCA GAGCGTGGGA 6240

CAGTGGCAAC CCTGGCATAG AACATTTGAT ACTCAGTTGC TCTAGGTCCT TGGCCTCCTT 6300

TCTTAGTCCT CCAAAACCAC AAACCCAGGG TTAAGGAAGC ATGGAATTAA TGTGAACAAA 6360

GCAACACCAT TGGTTTGGGC GATGAGACTG AGGCTTTTCT TCCTTTGTTT CTGTATTTTC 6420

TAGAATGCAG TAGTACCATG TATTACAGTA AAACAGCCAT ATTTTTGTGT CCTGTTCTGT 6480

AAAGGACAGA AGCCCCCATA TGCTTTGAGG GCAGTTTAGT TTATTAGAAG CAACAGAGCC 6540

TAGATTCAGC ACTGCCTGGT TTGGGACCTC CCTTTAGACA CCTCCCTTTT CTCACCTGTA 6600

AATAAAGGCT AAGTAAGCAT TTGTGACTGC ATACTCAGTC ATGGCCTGAA TCCTGGGAAC 6660

AAGGCAGCTA GCAGCTAGAG GCTGGAAAAC AGGACTGGAC CTCAGCAGCT CTACTGCATT 6720

ACTTCCCCTA GAAGCAGGGT GTGGCTACAC AAAACCAGAC AGATAATGTA TGGCTGAATG 6780

TAGATTCATG AAATGCTTGG AAAGACATTT ACTTATCAGT ATGTTTAATT CCCAAAATGG 6840

TCAGCAACAA TTCACACAAA ATTGATTATA AGTTTTTTCA ATTTGCTTAG CTGTTTAGTG 6900

TCCAGTAGAA ATAAGATTAC TATTCTATAA AGTGACAGAT GTTCATCTAG TTCCCATTGA 6960

TGGTGAAGAA CATTATGTCA TCCCAAAAGA TCGTTAACTT AGATCGTGGT TCTCTACCTT 7020

CCTGATGTTG TGTGACCCCC AACTGTGAAA TTATTTTCAT TGCTACTTCA CAACTATAAT 7080

TTTGCTTCTG TCATGAATCA TAAAGCAAAT ATCTGTGTTT TCTGATGGTC TTAGGTGACC 7140

CCTGTGAAAG GGTCATTTGA CTCTACCCCC TACATGGGTT GTGATCCACA GGTTGAGAAG 7200

CACTGACTTA GATTCTCAGA TTGCAAGTAG AGCAGCAGAA TTTCGAAGAA CAGCAGTGGC 7260

GACAGAAGCT GCTTTGGGCA GTTGTCATTT GTTAGCTTTC ATTGGCTCAT TTTGTATACA 7320

GATTTTCGGA AGTATTTCAG ACTTTATGTT ATGTAGCCTT TAGAGGCAAC AGTTCAGGAC 7380

TGGAGAGATG GCTCAAGGGT TAAGAGCACT GGCTGTTTTT TCAGAGGACC CATGTTTGAC 7440

TCACAGCACA CACATGGTGG CTCACAGCCA TCATGACTCC TGTTCCAAAG GATCTGATGT 7500

CTTCTTCTGA CCTCTGCAGA CACCAGGCAT GCATACATGC AGGCAAAATA CCCATCAATA 7560

TAAAAATAAA TAACTGGGAA ATATGCAAAT TCTTTAATAT GCAAATTCTT CTCTCCCCAA 7620

CTGCCATTTC CCATGCTCCA CCCTCATCCC TTCCCTCCTC TCTTACTTCT TTTGTTTGGA 7680

ATTCTTTAGA TAGCATCATC AAGGAGGCTC TGAGGCTTTC CAGTGCATCC TTGAATATCC 7740

GGACTGCTAA GGAGGATTTC ACTCTGCACC TTGAGGATGG CTCCTATAAC ATCCGAAAAG 7800

ACGACATCAT CGCTCTTTAT CCACAGTTAA TGCATTTGGA TCCTGCAATC TACCCAGACC 7860

CTCTGGTAAG TTTTTCTGCT CATCAAAGTT ATGTATCGAG GTGACAGTCA CCCAGGAATG 7920

TATTTGTAAT TACAGCTTTG ATTTGATCAT TAAAGTGAAG CCATAGGGAT TGTCCCTCTT 7980

TATTGCGGCA AATATTCATG TTTTGGAAAC TTTGGGTAGA GGCAAGAGTT TTGAACTTTT 8040

ACACCTAATA TTCATTTCAT AGTTTCTGCT AGACTATGTT TTCAGTCATA ACAAAACTAC 8100

CACCTTTTTT CCCCCTCACA AAGTACCCTC TCCCAAATTT ACACTAATGG AGGGTAATGC 8160

ATTTGACTTG ATCCTTAGAG TAGTTGTTTA GAGCCATTTT GCTTCTTTTG TCTAACTGAA 8220

GAATTAGTCT ACAGGTAGAA CAGGAGGTCC CTAGAGCTTC TTGGTCCACC AGCTCTTCAT 8280

AAGCTCTTTC CAGTATCACC TGGTTCAGTG CTTGGTGTTT GCTAACTTGT AGAGGATGGA 8340

TTTATTAGTA GAAAATTACT CTTTGGATCC TCCAGGTCAA GAAGGCAACA ACTTTCTATC 8400

ATAATAGCTC ATTGGCTTCT TGTCTCTTTG TTGCAGACTT TTAAATATGA TCGATACCTG 8460

GATGAGAACA AGAAGGCAAA GACCTCCTTC TATAGCAATG GAAACAAACT AAAGTATTTC 8520

TATATGCCAT TTGGATCCGG AGCTACAATA TGCCCTGGGA GACTATTTGC TGTCCAAGAA 8580

ATCAAGCAAT TTTTGATTCT GATGCTTTCA TACTTTGAAC TGGAGCTTGT GGAGAGTCAT 8640

GTCAAGTGTC CTCCTCTAGA CCAGTCCAGG GCAGGCTTGG GGATTTTGCC ACCATTAAAT 8700

GATATTGAGT TTAAATATAA ACTGAAACAT CTGTGACATG TGGTTGGAAG AAGAGGACAC 8760

TGGATGATGT TGCTGGACTG CAGCGAGTCT CACTAAACAA GCCCTTGGGA CAAATGCTCT 8820

CCTTTGCTTC CCAGCAACTG ACTGTGCCTA GGAAAAGAAC TGGTACCCCC GGCACCACTC 8880

TCTGTTCTCA CTGCCTGAGT TCCTGGGTGT TCAGATAGCT GAGGTCAGAG TTTCACCACT 8940

CTTAGAAGCA ATGTCTTTTG TTTTTATTTT CAAAATGAAG ATACTCCAAT TGGCAGATTT 9000

TTTTTCCTAA GGAAATTGCT TCATACTTTT ATGAAAACTG ATTAATTATG AAAAGGCTTC 9060

AAATTCACGT TTTAGTGAAA CTGTTATTTT TTTCACTAGT GAAGTTCTTC ATGTGTGAAC 9120

ATATACTATA AAAACATTTT AAGGGATCAT ATCATGCTTT GCATAAAGGG AAAGGAAAAT 9180

ATTATTCAAC TTTTTTTTTT GGTTTTTCTA GACAGGGTTT CTCTGTGTAG CTTTGGAGCC 9240

TATCCTGGCA CTCACTCTGT AGAGCAGGCT TGGTCTTGAA CTCACAGAGA TCTGCCTGCC 9300

TTTGCCTTCC GAGTGCTGGG ATTAAAGTCG TGCGTCACCA ATGCCTGGCT ATTTAACTTT 9360

TTCGATGTCT AGTGGTGAGA GCTTTGAAAA TGATGCTACT GTGTTGGGAA TACTATGGGA 9420

AATTTTGATG CTTCGCTGTT ACATTTAAAT TTATTGCTGC TGGAAATTGT CACCCCAGTT 9480

TTCAATTGCC CCTCTCTCTC CCTTTTAATA TTCACACTGA TGAGCAGAGT TTTTTAGAGA 9540

TTAAAAAGAC CTCCCCAGAG CCCTGTCTCT GATGTTTTTA AGCCTTTAAT CTCAGTACTC 9600

AGGAGGCAGA GGCAGGCAGA GCTCTGTGAG TTCGAGGCCA GCCTGATCTA CAGATCGAGT 9660

TCCAGGCAAG CCGGGGCTAC AGAATGAGAC CTTGTCACTA AAAGAAATAA ATAAGGTCAA 9720

TTTTATGTCA CAACTGATTA TGAATCATTG TAAAGGATAA ATTGAAAAAA AAGAACTCCA 9780

CGGGAATGAC CATTTAAATG GTCTATTTTA GCTAAAATTA ACTATGAATT ATGTGGAGTT 9840

CATTAAGTGT ATGTTGACGT TATATGTTCC TTTAAAATGT CTTATGTTTT ATCTCTGAAT 9900

GTCTTGTAGA TGGAGAGCAA TAATAGTGTT TAAATACTGA GTCAATAAGG TTTTATCTAT 9960

GTACTTTAAG AGCATTATTA GCTGTGTCAT TTTTACTGAT ATATCTAATA TATTTATATG 10020

TAAATTATAT TTATCTTTTA TCTTATACTA CAAATATAAG TAAATATTTT AAAACCAGTA 10080

ACTTTAAAAT TACCTACCTT TCAGAAATGA AAATAAGAAC ATTTGTGCTT TAACCTTTGA 10140

AATAGAATGT TTATTCATCC ACTGATAAGT TAAAATAATT TTATCTGATT TGTTTCAAGA 10200

AACTCAAAAA TATTCAAAGT AATCATGCAC TCAAAGGTCT TCGTAAGGTT ACAGAAAATT 10260

CAATAAAATC TTTTTTGTGT AGGGACTGAG TCAGGGTCTA GAAGATGCTT GGCAGGTACT 10320

CCAGTAGTGA GCTGGATCCA GAAGATTCCT TAAACTTTAA AATCTTAACA CTAAGTATTA 10380

TCACAGAGTT ATTACCTAAG TAGAATATTT TTCCTTTCCT TTTCAATTGA CAGAGTCCCA 10440

CAGCAACACA GCTGGCTGTA ACTCTTCACA TAGCTTGCGC AGGCTTTGAA CTCACTGTAC 10500

TCCTGCCTTT CCTTTTCTAG GAAATTATTT TCCACATCAA GAAAATTTAA TTGTTCCGAT 10560

GAGGTATAGA GTAACAAATT TCTGTTATAT ATTCATCTGT ATTAAACTGA ATTC 10614

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 44 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: YES

(vi) ORIGINAL SOURCE:
(A) ORGANISM: rat

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: Clontech, RL 102j

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 1..44
(D) OTHER INFORMATION: /note= "1. Cholesterol 7α-Hydroxylase; 2. promoter"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

TGTTTGCTTT GGTCACTCAA GTTCAAGTTA TTGGATCATG GTCC 44

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: YES

(vi) ORIGINAL SOURCE:
(A) ORGANISM: rat

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: Clontech, RL 102j

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 1..19
(D) OTHER INFORMATION: /note= "1. Cholesterol 7α-Hydroxylase; 2. Promoter"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

CTATGGACTT AGTTCAAGG 19

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: YES

(vi) ORIGINAL SOURCE:
(A) ORGANISM: rat

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: Clontech, RL 102j

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 1..18
(D) OTHER INFORMATION: /note= "1. Cholesterol 7α-Hydroxylase; 2. Promoter"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

TGTTCTGGAG CCTCTTCT 18

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 49 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: YES

(vi) ORIGINAL SOURCE:
(A) ORGANISM: rat

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: Clontech, RL 102j

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 1..49
(D) OTHER INFORMATION: /note= "1. Cholesterol 7α-Hydroxylase; 2. Promoter"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

TCACTGTGGC CTAGTGCCAC ATCTACCTAT TTCTTTGGCT TTACTTTGT 49

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: YES

(vi) ORIGINAL SOURCE:
(A) ORGANISM: rat

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: Clontech, RL 102j

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 1..12
(D) OTHER INFORMATION: /note= "1. Cholesterol 7α-Hydroxylase; 2. Promoter"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

TGGTCAAGTT CA 12

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 126 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: YES

(vi) ORIGINAL SOURCE:
(A) ORGANISM: rat

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: Clontech, RL 102j

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 1..126
(D) OTHER INFORMATION: /note= "1. Cholesterol 7α-Hydroxylase; 2. Promoter"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

CTAGTAGGAG GACAAATAGT GTTTGCTTTG GTCACTCAAG TTCAAGTTAT TGGATCATGG 60

TCC 63

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 70 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: YES

(vi) ORIGINAL SOURCE:
(A) ORGANISM: rat

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: Clontech, RL 102j

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 1..70
(D) OTHER INFORMATION: /note= "1. Cholesterol 7α-Hydroxylase; 2. Promoter"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

CCTCTTCTGA GACTATGGAC TTAGTTCAAG GCCGG 35

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 120 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: YES

(vi) ORIGINAL SOURCE:
(A) ORGANISM: rat

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: Clontech, RL 102j

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 1..120
(D) OTHER INFORMATION: /note= "1. Cholesterol 7α-Hydroxylase; 2. Promoter"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

TCACTGTGGC CTAGTGCCAC ATCTACCTAT TTCTTTGGCT TTACTTTGTG CTAGGTGACC 60

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 43 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: YES

(vi) ORIGINAL SOURCE:
(A) ORGANISM: rat

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: Clontech, RL 102j

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 1..43
(D) OTHER INFORMATION: /note= "1. Cholesterol 7α-Hydroxylase; 2. Promoter"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

GAAGATCTAG TAGGAGGACA AATAG 25

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 45 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: YES

(vi) ORIGINAL SOURCE:
(A) ORGANISM: rat

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: Clontech, RL 102j

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 1..45
(D) OTHER INFORMATION: /note= "1. Cholesterol 7α-Hydroxylase; 2. Promoter"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

GATCCTTGGT CACTCAAGTT C 21

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: YES

(vi) ORIGINAL SOURCE:
(A) ORGANISM: rat

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: Clontech, RL 102j

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 1..39
(D) OTHER INFORMATION: /note= "1. Cholesterol 7α-Hydroxylase; 2. Promoter"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

GATCCAATAG TGTTTGCTTT GGT 23

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: YES

(vi) ORIGINAL SOURCE:
(A) ORGANISM: rat

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: Clontech, RL 102j

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 1..29
(D) OTHER INFORMATION: /note= "1. Cholesterol 7α-Hydroxylase; 2. Promoter"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

AGATGGCTCG AGACTCTTTG CCTAGCAAA 29

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: YES

(vi) ORIGINAL SOURCE:
(A) ORGANISM: rat

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: Clontech, RL 102j

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 1..17
(D) OTHER INFORMATION: /note= "1. Cholesterol 7α-Hydroxylase; 2. Promoter"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

CAGCACATGA GGGACAG 17

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: YES

(vi) ORIGINAL SOURCE:
(A) ORGANISM: rat

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: Clontech, RL 102j

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 1..19
(D) OTHER INFORMATION: /note= "1. Cholesterol 7α-Hydroxylase; 2. Promoter"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

CTCTTCTGAG ACTATGGAC 19

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: YES

(vi) ORIGINAL SOURCE:
(A) ORGANISM: rat

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: Clontech, RL 102j

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 1..25
(D) OTHER INFORMATION: /note= "1. Cholesterol 7α-Hydroxylase; 2. Promoter"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

GAAGATCTAG TAGGAGGACA AATAG 25

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 264 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: YES

(vi) ORIGINAL SOURCE:
(A) ORGANISM: rat

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: Clontech, RL 102j

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 1..264
(D) OTHER INFORMATION: /note= "1. Cholesterol 7α-Hydroxylase; 2. Promoter"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

CAGCACATGA GGGACAGACC TTCAGCTTAT CGAGTATTGC AGCTCTCTGT TTGTTCTGGA 60

GCCTCTTCTG AGACTATGGA CTTAGTTCAA GGCCGGGTAA TGCTATTTTT TTCTTCTTTT 120

TTCTAGTAGG AGGAGGACAA ATAGTGTTTG CTTTGGTCAC TCAAGTTCAA GTTATTGGAT 180

CATGGTCCTG TGCACATATA AAGTCTAGTC AGACCCACTG TTTCGGGACA GCCTTGCTTT 240

GCTAGGCAGG CAAAGAGTCT CGAG 264

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 199 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: YES

(vi) ORIGINAL SOURCE:
(A) ORGANISM: rat

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: Clontech, RL 102j

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 1..199
(D) OTHER INFORMATION: /note= "1. Cholesterol 7α-Hydroxylase; 2. Promoter"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

CTCTTCTGAG ACTATGGACT TAGTTCAAGG CCGGGTAATG CTATTTTTTT CTTCTTTTTT 60

CTAGTAGGAG GACAAATAGT GTTTGCTTTG GTCACTCAAG TTCAAGTTAT TGGATCATGG 120

TCCTGTGCAC ATATAAAGTC TAGTCAGACC CACTGTTTCG GGACAGCCTT GCTTTGCTAG 180

GCAGGCAAAG AGTCTCGAG 199

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 145 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: YES

(vi) ORIGINAL SOURCE:
(A) ORGANISM: rat

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: Clontech, RL 102j

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 1..145
(D) OTHER INFORMATION: /note= "1. Cholesterol 7α-Hydroxylase; 2. Promoter"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

GAAGATCTAG TAGGAGGACA AATAGTGTTT GCTTTGGTCA CTCAAGTTCA AGTTATTGGA 60

TCATGGTCCT GTGCACATAT AAAGTCTAGT CAGACCCACT GTTTCGGGAC AGCCTTGCTT 120

TGCTAGGCAG GCAAAGAGTC TCGAG 145

(2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 86 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iii) ANTI-SENSE: YES

(vi) ORIGINAL SOURCE:
(A) ORGANISM: rat

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: Clontech, RL 102j

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 1..86
(D) OTHER INFORMATION: /note= "1. Cholesterol 7α-Hydroxylase; 2. Promoter"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

GAAGATCTAG TAGGAGGACA AATAGTGTTT GATTTGGTCA CTCAAGTTCA AGTTATTGGA 60

TCATGGTCCT GTGCACATCC TAGGGC 86

Claims

1. A regulatory element of the cholesterol 7α-hydroxylase (CYP7) gene selected from DNA fragments in the group consisting of from about -160 to about +32, from about -3643 to about -224, from about -224 and +32, from about -191 to about +64 of the rat CYP7 gene, from about -252 to about +3 of the hamster CYP7 gene, and from about -187 to about +65, from about -158 to about +32, from about -3643 to about -224, from about -223 to about +32, of the human CYP7 gene.

2. A regulatory element of the rat CYP7 gene selected from DNA fragments in the group consisting of from about -101 to about -29, from about -81 to about -37, from about -161 to about -127, from about -149 to about -131, from about -171 to about -154, from about -101 to about -82, from about -73 to about -56, and from about -86 to about -71.

3. A regulatory element of the human CYP7 gene selected from DNA fragments in the group consisting of from about -104 to about -30, from about -78 to about -36, from about -159 to about -124, from about -147 to about -128, from about -169 to about -152, from about -104 to about -79, from about -71 to about -54 and from about -89 to about -68.

4. A regulatory element of hamster CYP7 gene selected from DNA fragments in the group consisting of from about -161 to about -86, from about -136 to about -92, from about -208 to about -184, from about -206 to about -188, from about -228 to about -211, from about -161 to about -137, from about -128 to about -111 and from about -146 to about -126.

5. A construct comprising at least one regulatory element as defined in claim 1, wherein said regulatory element is operably attached to a structural gene.

6. A construct according to claim 5, wherein said structural gene is a reporter gene.

7. A construct according to claim 6, wherein said structural gene comprises the gene encoding luciferase.

8. A host cell transformed with a vector comprising a construct according to claim 6.

9. A host cell according to claim 8 that is a HepG2 cell.

10. A host cell according to claim 9 that is a confluent HepG2 cell.

11. A method for determining whether an agent inhibits or stimulates CYP7 gene expression comprising the steps of:
(a) providing a host cell according to claim 9 in a medium suitable for expression of said structural gene;
(b) contacting said host cell with said agent; and
(c) detecting an inhibition or stimulation of gene expression.

12. A method according to claim 9, wherein said agent is a physiological agent endogenous to a human.

13. A method according to claim 11, wherein said agent is an agent exogenous to a human.

14. A method for detecting a transcription factor of CYP7, comprising the step of contacting a fragment of DNA according to claim 1 with a biological sample suspected of containing a transcription factor and detecting binding between said fragment and a transcription factor.

15. A method for detecting a transcription factor according to claim 14, wherein said binding is detected by performing a footprint analysis.

16. A method according to claim 14 further comprising the step of isolating the transcription factor.

17. A substantially isolated CYP7 transcription factor identified by the process of claim 16, wherein the factor binds to a core sequence comprising (T or C)CAAG(T or C).

18. A transcription factor according to claim 17 wherein said factor binds to a sequence comprising TCAAGTTCAAGT or CCAAGCTCAAGT.

19. A transcription factor according to claim 17 that is characterized by a molecular weight of about 57,000 Daltons.

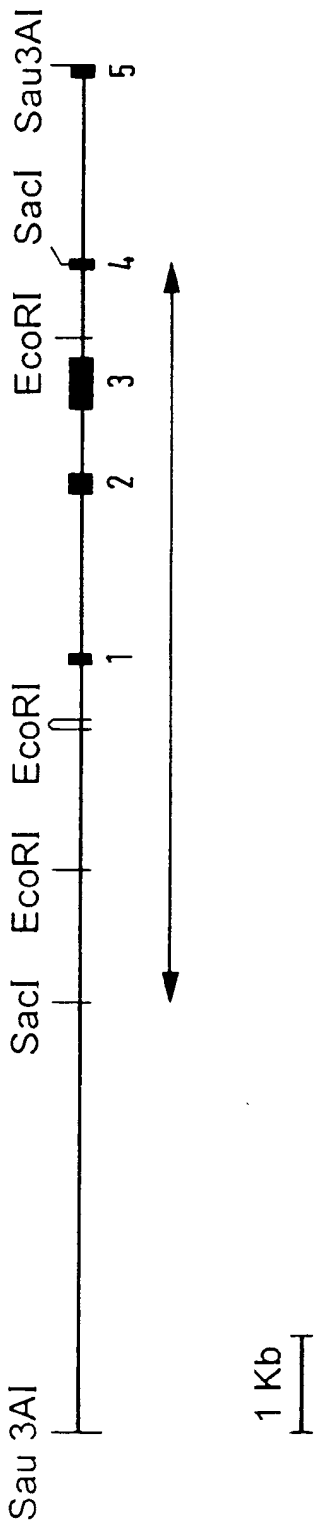


Fig. 1

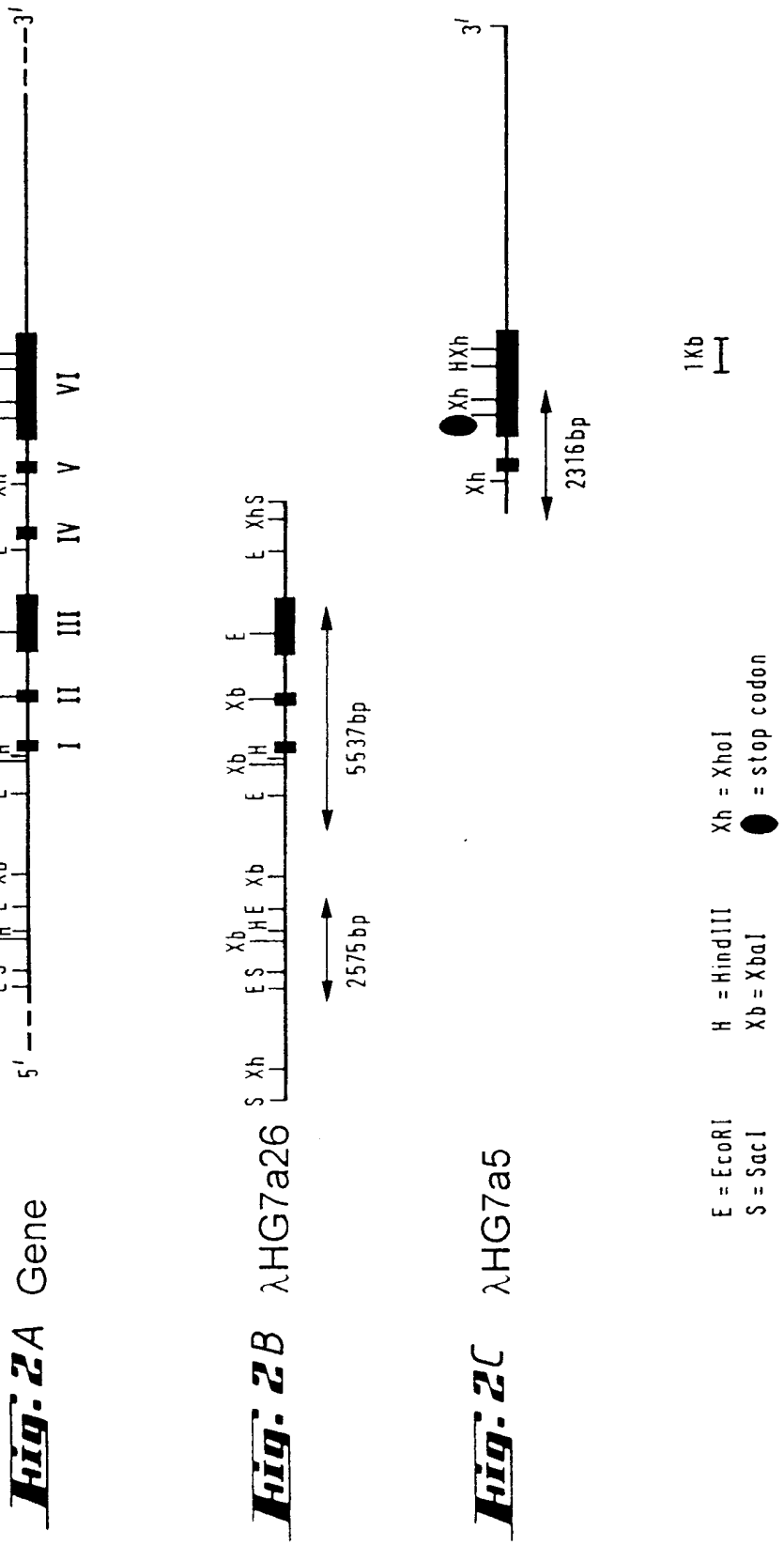


Fig. 2A

Fig. 2B

Fig. 2C

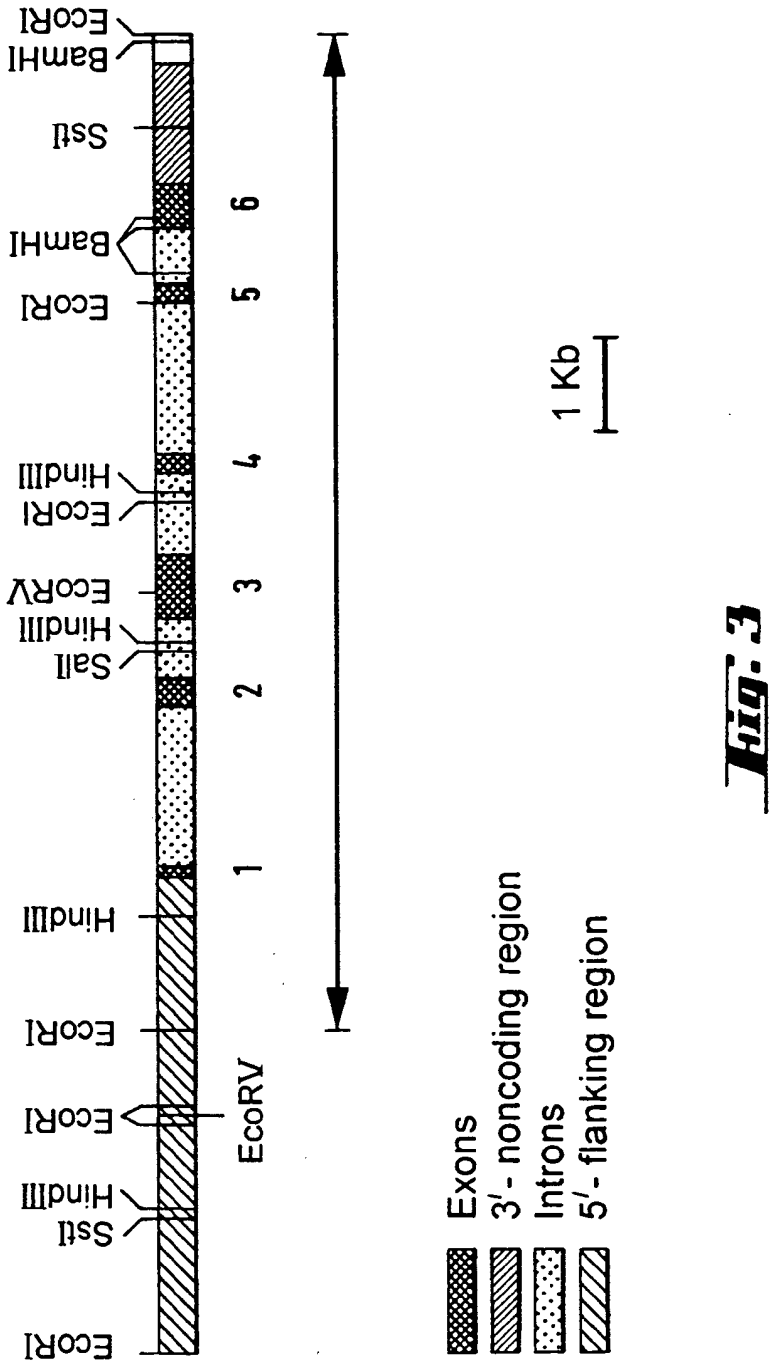


Fig. 3

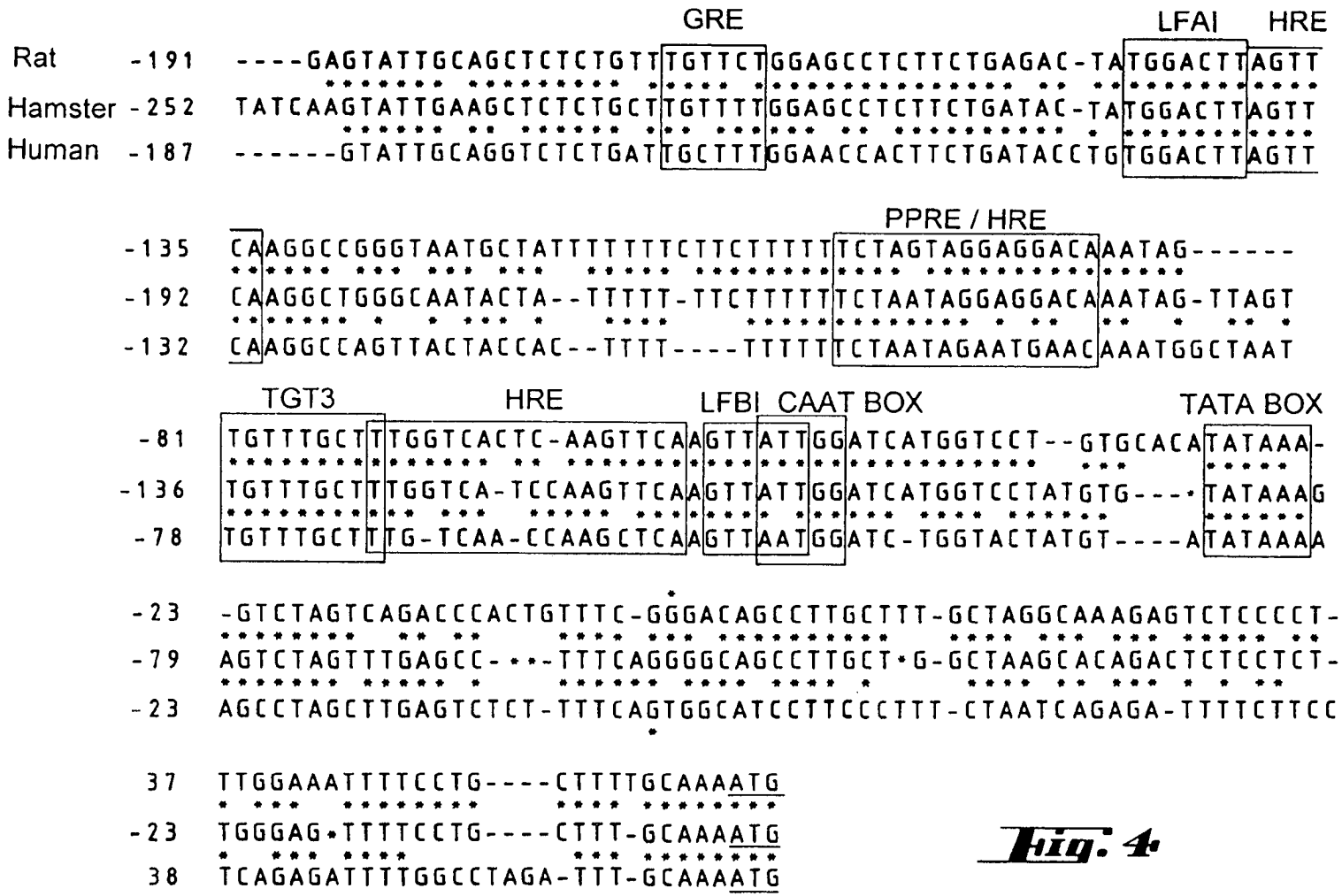
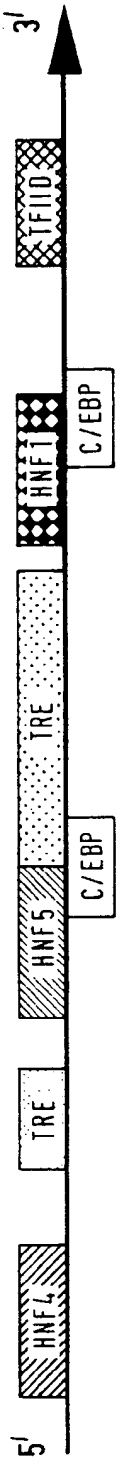


Fig. 4

Fig. 5



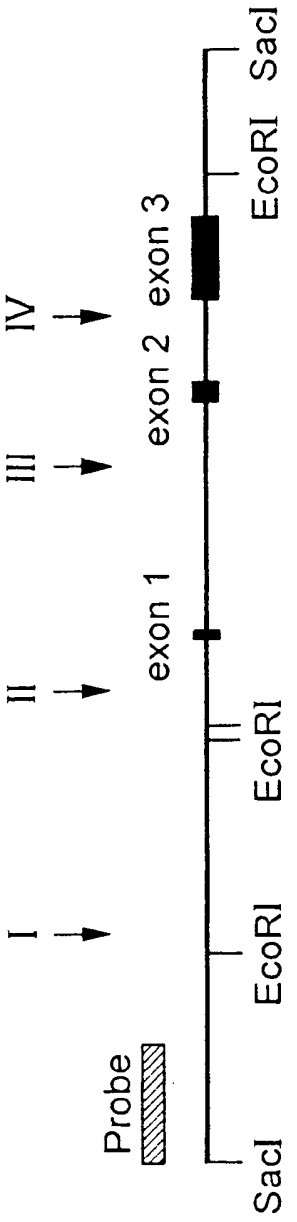


Fig. 6

Fig. 1

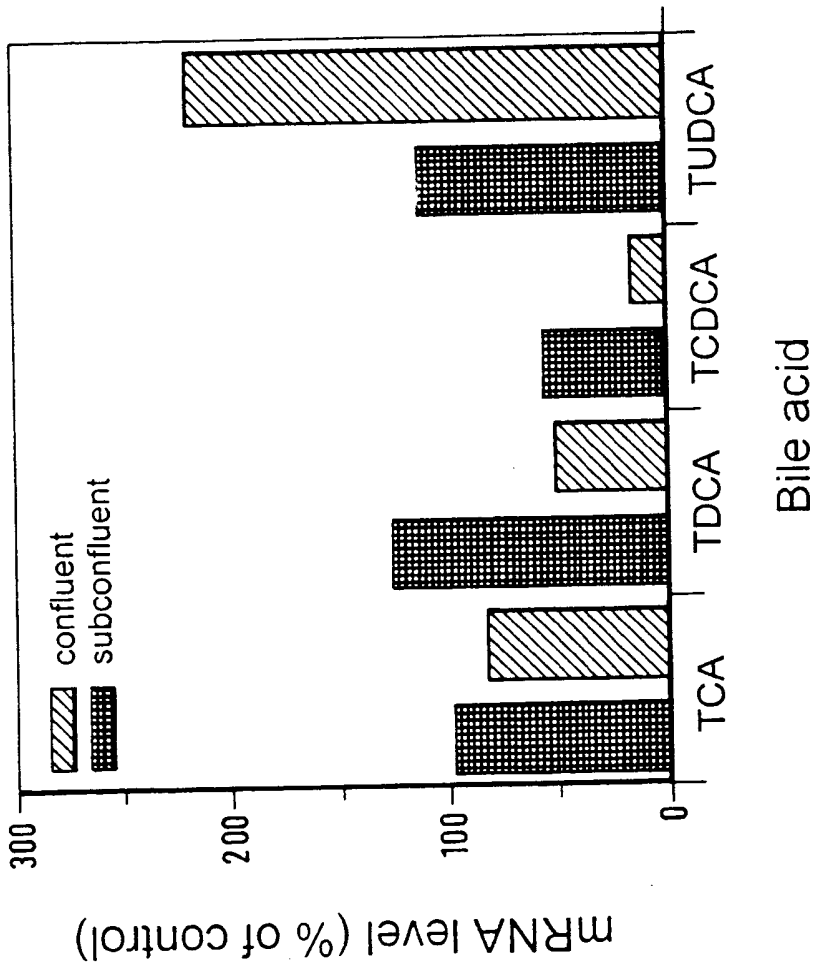
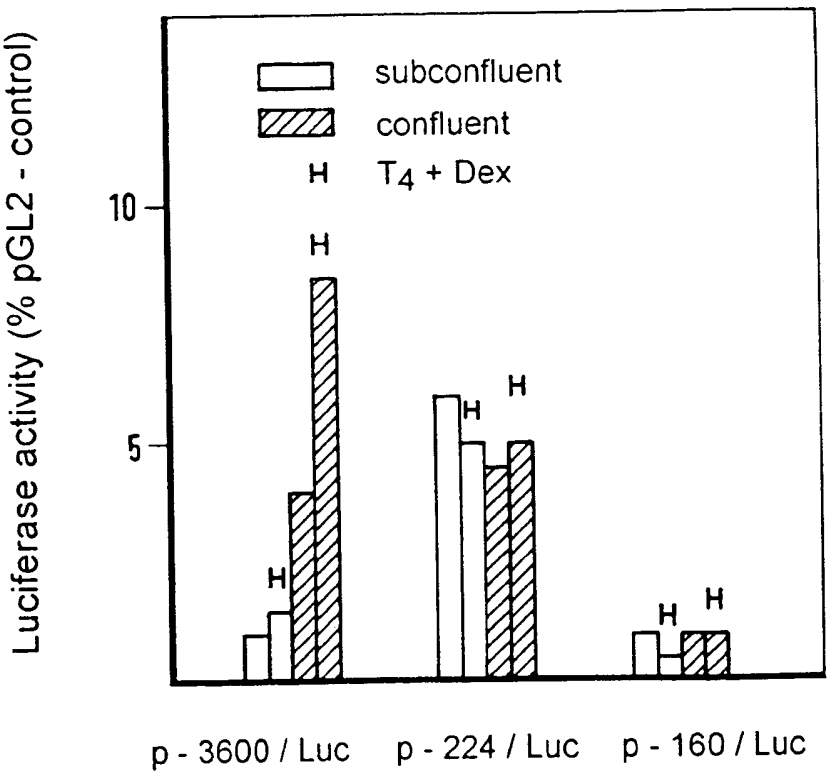


Fig. 8

Effect of T_4 and Dexamethasone on transcriptional activity of CYP7 / Luc constructs in HepG2 cell cultures



(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 0 648 840 A3

(12) EUROPEAN PATENT APPLICATION

(88) Date of publication A3:
20.08.1997 Bulletin 1997/34

(43) Date of publication A2:
19.04.1995 Bulletin 1995/16

(21) Application number: **94115856.0**

(22) Date of filing: **07.10.1994**

(51) Int. Cl.[6]: **C12N 15/53**, C12N 9/02, C07K 14/47, C12Q 1/68, C12Q 1/26, C12N 5/10, C12N 15/85

(84) Designated Contracting States:
AT BE CH DE DK ES FR GB GR IE IT LI LU MC NL PT SE

(30) Priority: **13.10.1993 US 135511**
13.10.1993 US 135488
13.10.1993 US 135510
28.01.1994 US 187453

(71) Applicant: **Northeastern Ohio Universities**
Rootstown, Ohio 44272-0095 (US)

(72) Inventor: **Chiang, John Young Ling**
Stow, Ohio 44224 (US)

(74) Representative: **Bösl, Raphael Konrad, Dr.**
Hoechst AG
Zentrale Patentabteilung
Gebäude F 821
65926 Frankfurt am Main (DE)

(54) Cholesterol 7alpha-hydroxylase gene regulatory elements and transcription factors

(57) DNA regulatory elements that control cholesterol 7α-hydroxylase expression are disclosed, including bile acid responsive elements. A gene construct comprising at least one CYP7 regulatory element and a reporter gene is used to transfect HepG2 cells. Confluent transfected HepG2 cells are employed in an assay to detect a compound that modulates cholesterol 7α-hydroxylase enzyme regulation. A method for screening compounds that inhibit or stimulate expression of the enzyme is provided, as well as a method for detecting and isolating transcription factors of the cholesterol 7α-hydroxylase gene. A transcription factor of 57 KDa is identified which is useful in an assay for determining regulation of CYP7 expression.

Fig. 5

EP 0 648 840 A3

Printed by Rank Xerox (UK) Business Services
2.14.12/3.4

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 94 11 5856

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.6)
X	WO 92 18523 A (MERCK & CO INC ;MOLOWA DAVID T (US)) 29 October 1992 * the whole document * ---	1,3,5,6, 8-10, 14-16	C12N15/53 C12N9/02 C07K14/47 C12Q1/68 C12Q1/26
X	GENE, vol. 130, 25 August 1993, pages 217-223, XP002026711 HOEKMAN, M.F.M., ET AL . : "TRANSCRIPTIONAL REGULATION OF THE GENE ENCODING CHOLESTEROL 7ALPHA-HYDROXYLASE IN THE RAT" * the whole document * ---	1,2,5,6, 8	C12N5/10 C12N15/85
X,D	BIOCHIMICA ET BIOPHYSICA ACTA, vol. 1132, 1992, pages 337-339, XP000674487 CHIANG, J.Y.L., ET AL . : "CLONING AND 5'-FLANKING SEQUENCE OF A RAT CHOLESTEROL 7ALPHA-HYDROXYLASE GENE" * the whole document * ---	1,2	
X	BIOCHIMICA ET BIOPHYSICA ACTA, vol. 1168, 1993, pages 239-242, XP000674632 THOMPSON, J.F., ET AL . : "CHOLESTEROL 7ALPHA-HYDROXYLASE PROMOTER SEPARATED FROM CYCLOPHILIN PSEUDOGENE BY ALU SEQUENCE" * the whole document * ---	1,3	TECHNICAL FIELDS SEARCHED (Int.Cl.6) C07K C12N C12Q
X,D	BIOCHIMICA ET BIOPHYSICA ACTA, vol. 1172, 1993, pages 147-150, XP000674597 NISHIMOTO,M., ET AL . : "STRUCTURE OF THE GENE ENCODING HUMAN LIVER CHOLESTEROL 7ALPHA-HYDROXYLASE" * the whole document * ---	1,3	
Y	-/--	5-10	

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	11 June 1997	Holtorf, S

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 94 11 5856

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.6)
X	EMBL SEQUENCE DATA LIBRARY, HEIDELBERG, GERMANY, XP002026713 CRESTANI,M., ET AL . : "GENOMIC CLONING, SEQUENCING, AND ANALYSIS OF THE HAMSTER CHOLESTEROL 7ALPHA-HYDROXYLASE GENE (CYP7)" ACCESSION NO. L04690 ---	1,4	
P,X P,Y	WO 94 18346 A (UNIV TEXAS) 18 August 1994 * the whole document * ---	14-16 5-10	
P,X	THE JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 269, no. 26, 1 July 1994, pages 17502-17507, XP002032493 CHIANG, J.Y.L., ET AL . : "IDENTIFICATION AND CHARACTERIZATION OF A PUTATIVE BILE ACID-RESPONSIVE ELEMENT IN CHOLESTEROL 7alpha-HYDROXYLASE GENE PROMOTER" * the whole document * ---	1,2, 5-12,14, 15,17-19	
P,X	THE JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 269, no. 20, 20 May 1994, pages 14681-14689, XP002032494 LEE, Y-H., ET AL . : "MULTIPLE , FUNCTIONAL DBP SITES ON THE PROMOTER OF THE CHOLESTEROL 7alpha-HYDROXYLASE P450 GENE, CYP7" * the whole document * ---	1,2, 5-10,14, 15	TECHNICAL FIELDS SEARCHED (Int.Cl.6)
	-/--		

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	11 June 1997	Holtorf, S

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 94 11 5856

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.6)
P,X	BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS, vol. 198, no. 2, 28 January 1994, pages 546-553, XP002032495 CRESTANI, M., ET AL . : "EFFECTS OF BILE ACIDS AND STEROID/THYROID HORMONES ON THE EXPRESSION OF CHOLESTEROL 7 alpha-HYDROXYLASE mRNA AND THE CYP7 GENE IN HepG2 CELLS" * the whole document * ---	1,2,5-13	
T	THE JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 272, no. 6, - 7 February 1997 pages 3444-3452, XP002032496 COOPER, A.D., ET AL.: "CHARACTERIZATION OF HEPATIC-SPECIFIC REGULATORY ELEMENTS IN THE PROMOTER REGION OF THE HUMAN CHOLESTEROL 7alpha-HYDROXYLASE GENE" * the whole document * -----	1-19	
			TECHNICAL FIELDS SEARCHED (Int.Cl.6)

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	11 June 1997	Holtorf, S

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

European Patent Office

CLAIMS INCURRING FEES

The present European patent application comprised at the time of filing more than ten claims.

☐ All claims fees have been paid within the prescribed time limit. The present European search report has been drawn up for all claims.

☐ Only part of the claims fees have been paid within the prescribed time limit. The present European search report has been drawn up for the first ten claims and for those claims for which claims fees have been paid,

namely claims:

☐ No claims fees have been paid within the prescribed time limit. The present European search report has been drawn up for the first ten claims.

LACK OF UNITY OF INVENTION

The Search Division considers that the present European patent application does not comply with the requirement of unity of invention and relates to several inventions or groups of inventions,
namely:

see sheet -B-

☒ All further search fees have been paid within the fixed time limit. The present European search report has been drawn up for all claims.

☐ Only part of the further search fees have been paid within the fixed time limit. The present European search report has been drawn up for those parts of the European patent application which relate to the inventions in respects of which search fees have been paid,

namely claims:

☐ None of the further search fees has been paid within the fixed time limit. The present European search report has been drawn up for those parts of the European patent application which relate to the invention first mentioned in the claims,

namely claims:

European Patent Office

EP 94 11 5856 - B -

LACK OF UNITY OF INVENTION

The Search Division considers that the present European patent application does not comply with the requirement of unity of invention and relates to several inventions, or groups of inventions, namely:

1. Claims 1-13:
   Use of rat, hamster and human Promoter-sequences of the CYP7 gene in a screening method including recombinant cells to detect agents that stimulate or inhibit CYP7 gene expression; regulatory element for use in said assay

2. Claims 14-19:
   Method of detection of CYP7 Transcription factors as determined by a footprint-analysis and transcription factors identified by said method

EPO Form Supplementary Sheet B (1996)